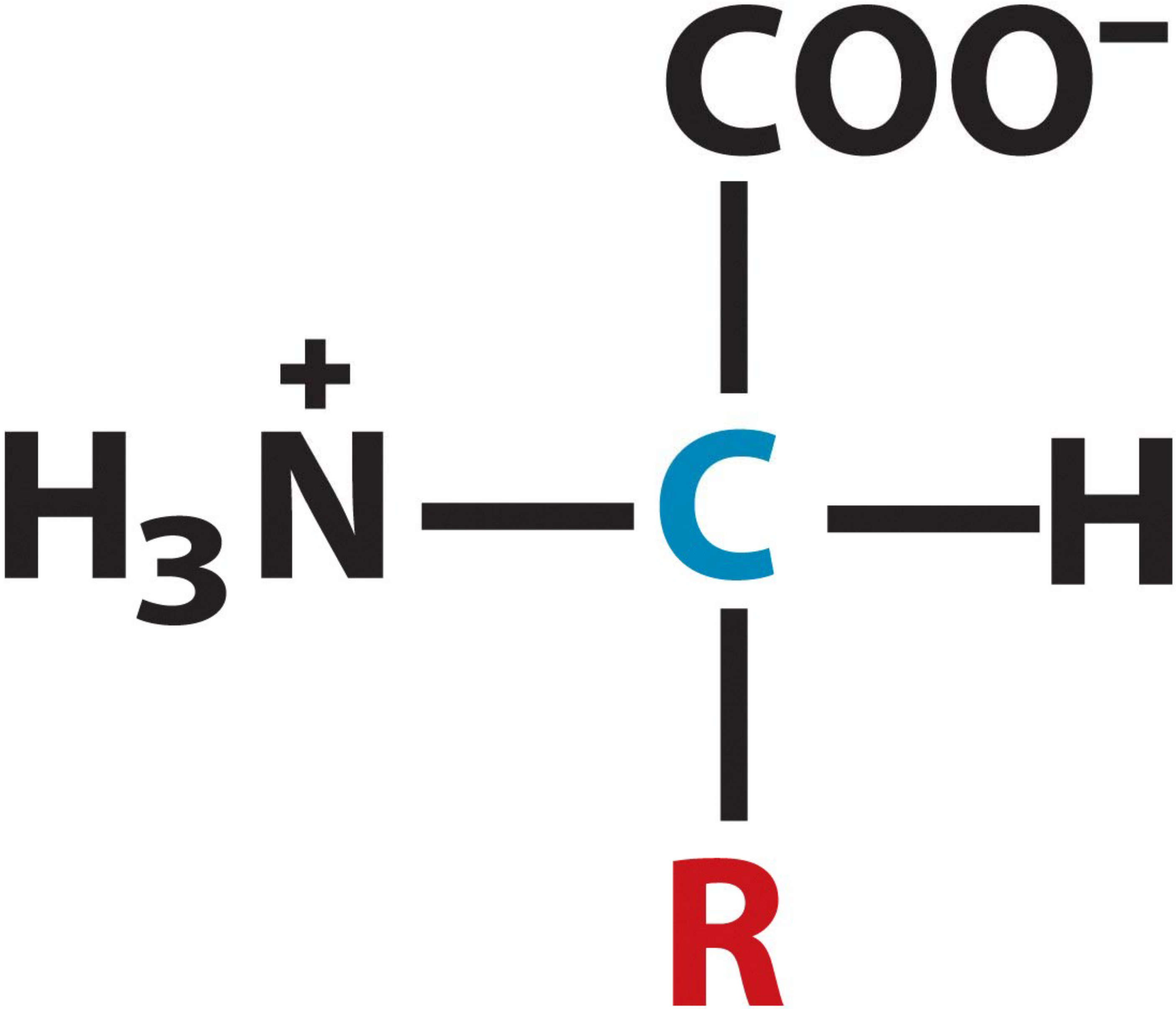
David L. Nelson and Michael M. Cox

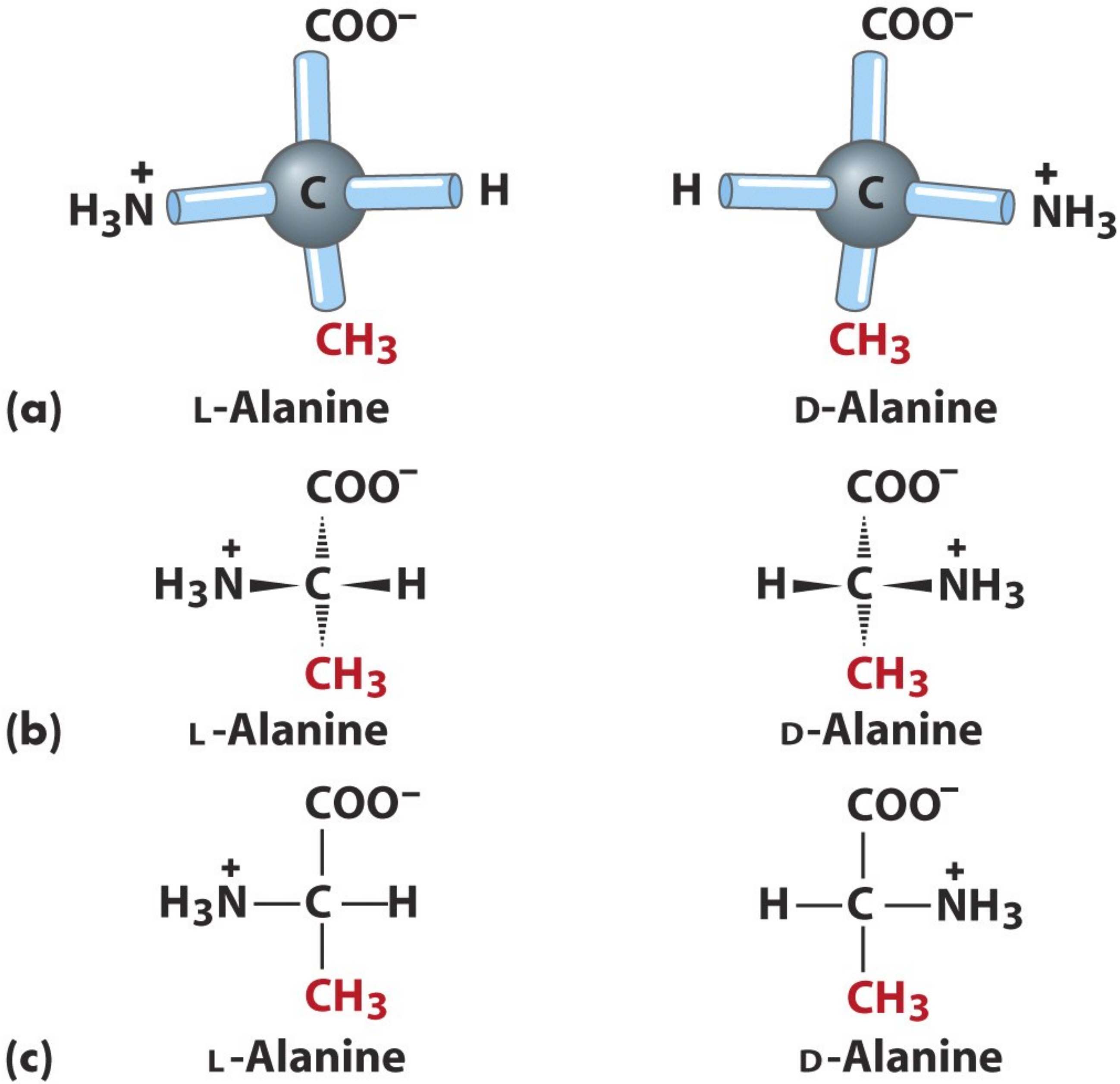
Lehninger Principles of Biochemistry

Fourth Edition

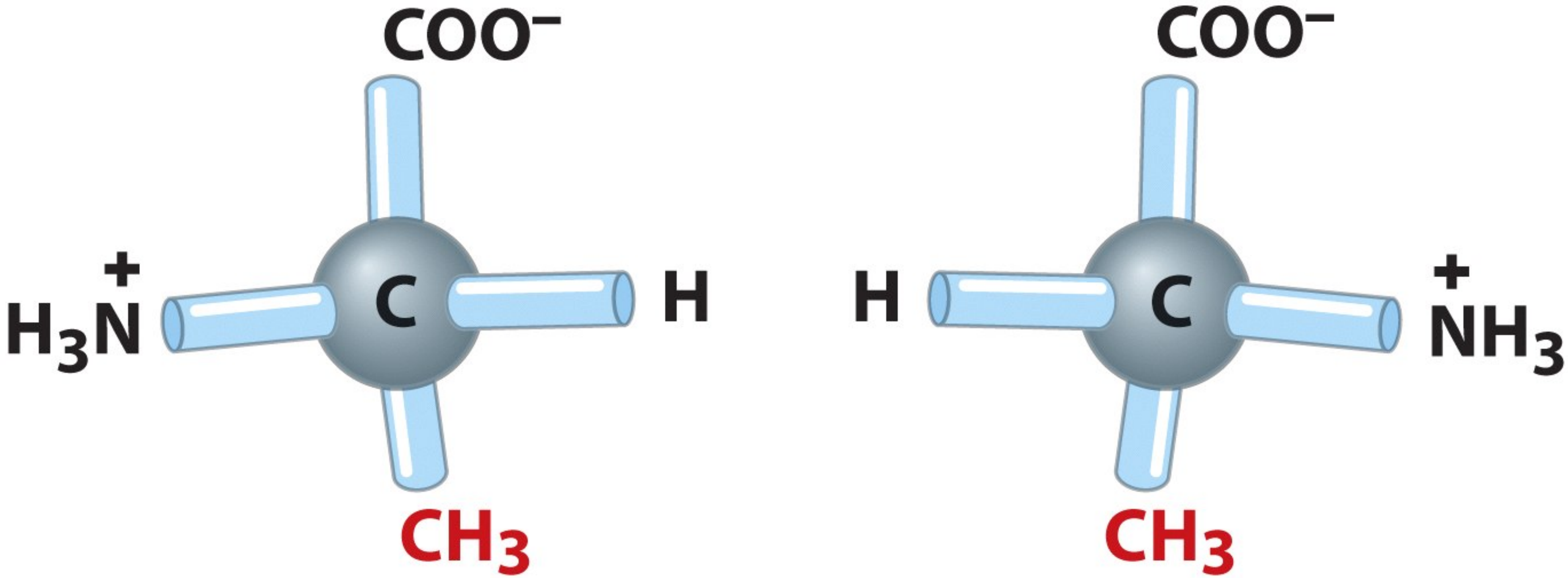
Chapter 3:

Amino Acids, Peptides, and Proteins

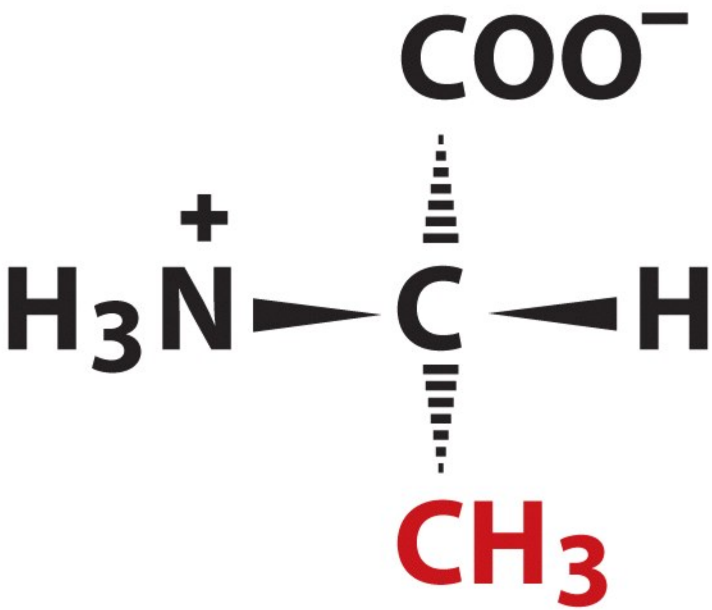




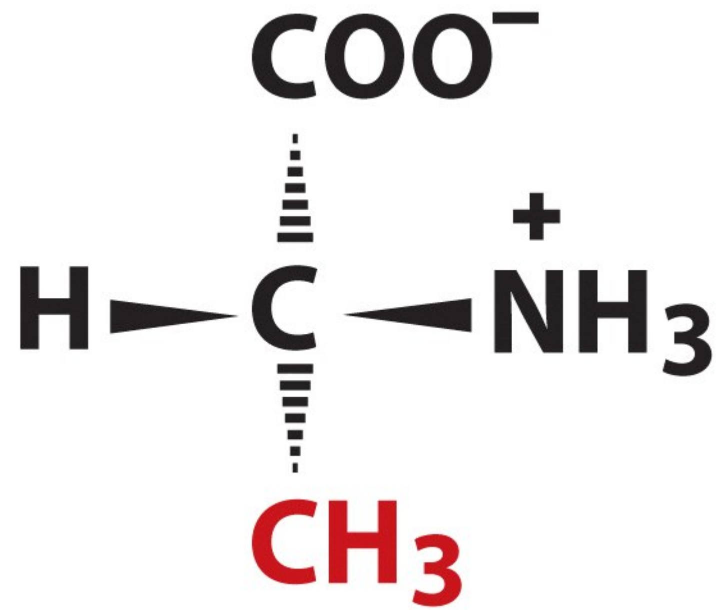
COO⁻
H₃N⁺
C
H
CH₃
(a)
L-Alanine
D-Alanine
NH₃⁺
(b)
(c)



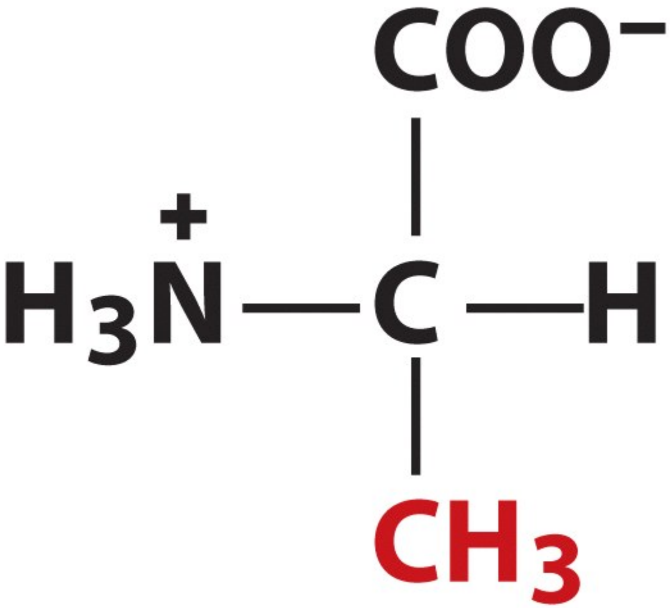
L-Alanine D-Alanine



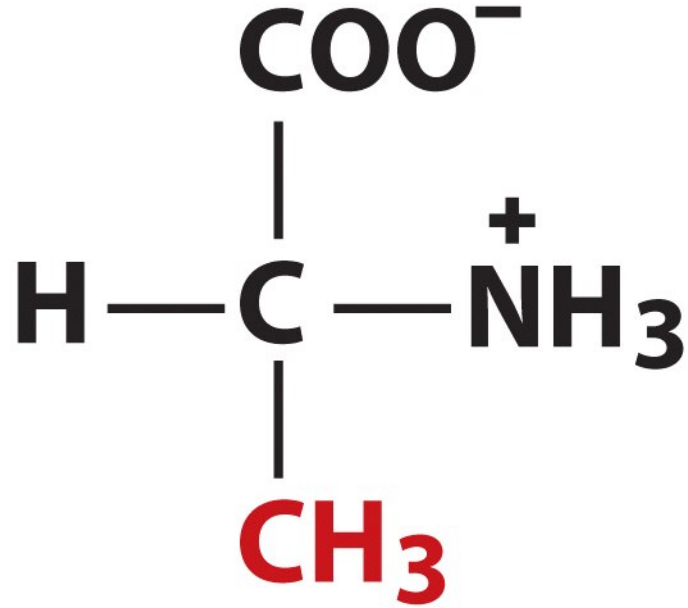
L-Alanine



D-Alanine



L-Alanine



D-Alanine

TABLE 3–1 Properties and Conventions Associated with the Common Amino Acids Found in Proteins

Amino acid	Abbreviation/ symbol	M_r	pK_a values: pK_1 (—COOH)	pK_2 ($—NH_3^+$)	pK_R (R group)	pI	Hydropathy index*	Occurrence in proteins (%)†
Nonpolar, aliphatic R groups								
Glycine	Gly G	75	2.34	9.60		5.97	−0.4	7.2
Alanine	Ala A	89	2.34	9.69		6.01	1.8	7.8
Proline	Pro P	115	1.99	10.96		6.48	1.6	5.2
Valine	Val V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu L	131	2.36	9.60		5.98	3.8	9.1
Isoleucine	Ile I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups								
Phenylalanine	Phe F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr Y	181	2.20	9.11	10.07	5.66	−1.3	3.2
Tryptophan	Trp W	204	2.38	9.39		5.89	−0.9	1.4

*A scale combining hydrophobicity and hydrophilicity of R groups; it can be used to measure the tendency of an amino acid to seek an aqueous environment (− values) or a hydrophobic environment (+ values). See Chapter 11. From Kyte, J. & Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132.

†Average occurrence in more than 1,150 proteins. From Doolittle, R.F. (1989) Redundancies in protein sequences. In *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G.D., ed.), pp. 599–623, Plenum Press, New York.

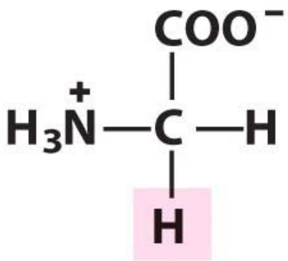
TABLE 3–1 Properties and Conventions Associated with the Common Amino Acids Found in Proteins

Amino acid	Abbreviation/ symbol	M_r	pK_a values: pK_1 (—COOH)	pK_2 ($—NH_3^+$)	pK_R (R group)	pI	Hydropathy index*	Occurrence in proteins (%)†
Polar, uncharged R groups								
Serine	Ser S	105	2.21	9.15		5.68	−0.8	6.8
Threonine	Thr T	119	2.11	9.62		5.87	−0.7	5.9
Cysteine	Cys C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn N	132	2.02	8.80		5.41	−3.5	4.3
Glutamine	Gln Q	146	2.17	9.13		5.65	−3.5	4.2
Positively charged R groups								
Lysine	Lys K	146	2.18	8.95	10.53	9.74	−3.9	5.9
Histidine	His H	155	1.82	9.17	6.00	7.59	−3.2	2.3
Arginine	Arg R	174	2.17	9.04	12.48	10.76	−4.5	5.1
Negatively charged R groups								
Aspartate	Asp D	133	1.88	9.60	3.65	2.77	−3.5	5.3
Glutamate	Glu E	147	2.19	9.67	4.25	3.22	−3.5	6.3

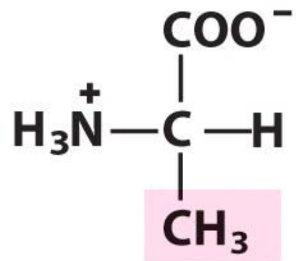
*A scale combining hydrophobicity and hydrophilicity of R groups; it can be used to measure the tendency of an amino acid to seek an aqueous environment (− values) or a hydrophobic environment (+ values). See Chapter 11. From Kyte, J. & Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132.

†Average occurrence in more than 1,150 proteins. From Doolittle, R.F. (1989) Redundancies in protein sequences. In *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G.D., ed.), pp. 599–623, Plenum Press, New York.

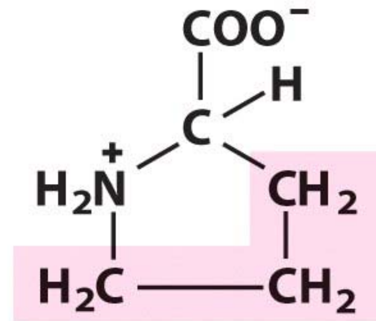
Nonpolar, aliphatic R groups



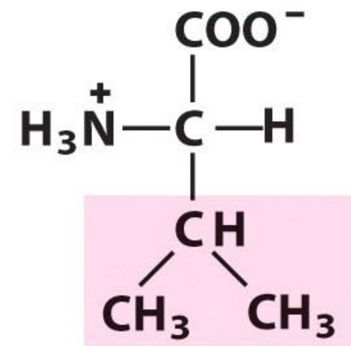
Glycine



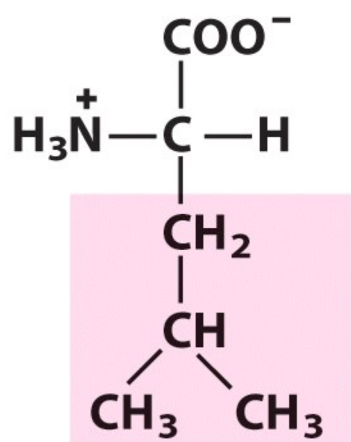
Alanine



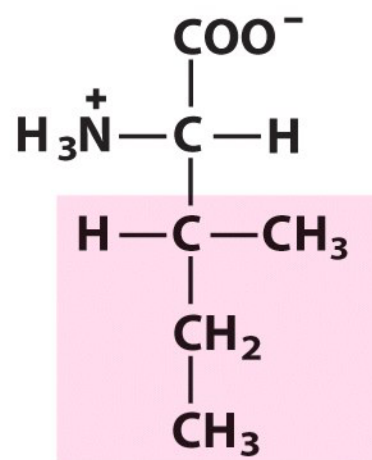
Proline



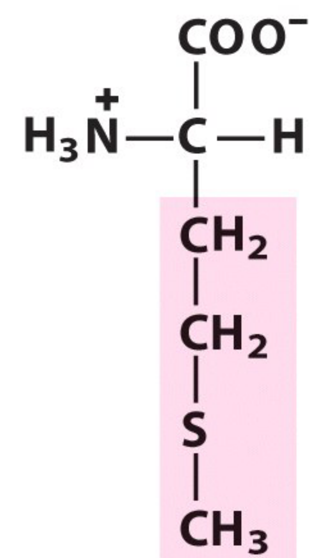
Valine



Leucine

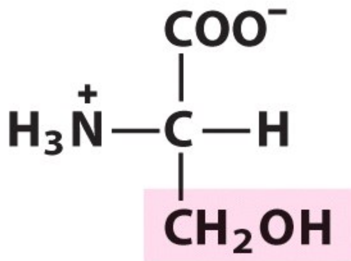


Isoleucine

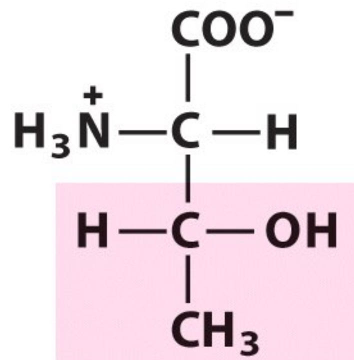


Methionine

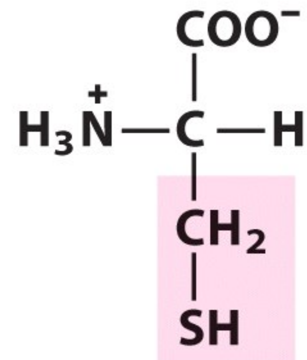
Polar, uncharged R groups



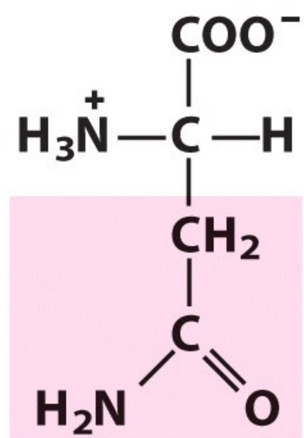
Serine



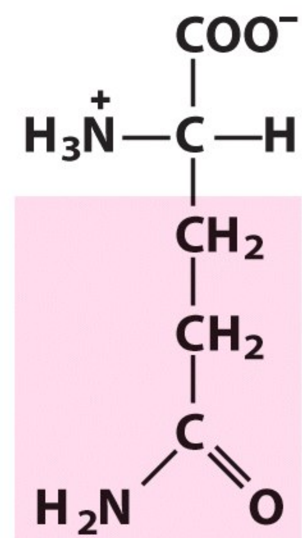
Threonine



Cysteine

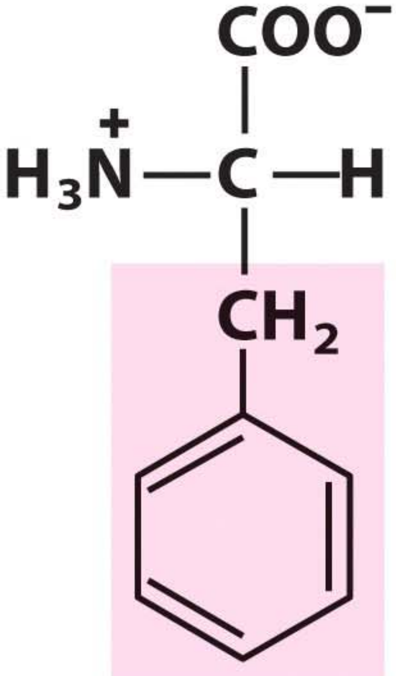


Asparagine

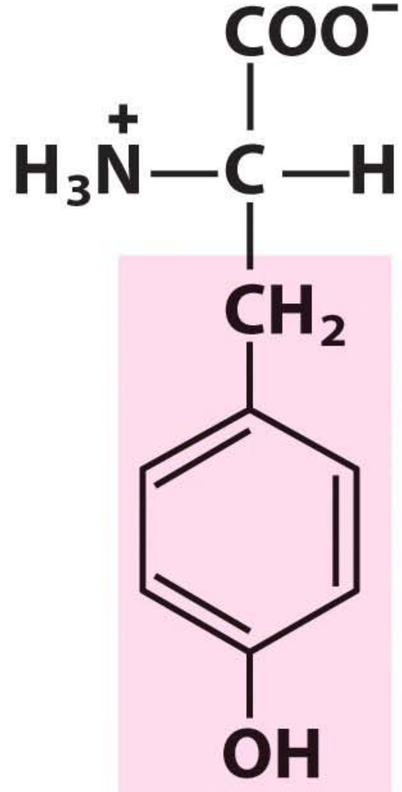


Glutamine

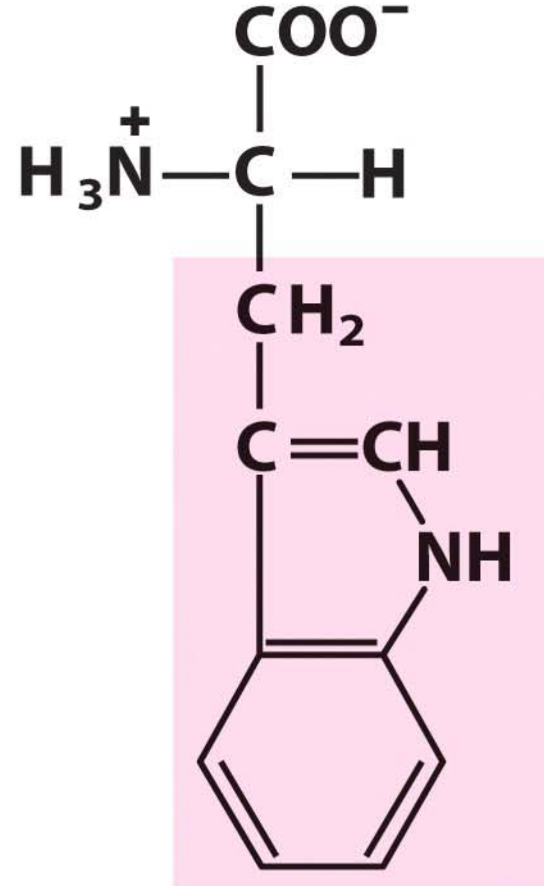
Aromatic R groups



Phenylalanine

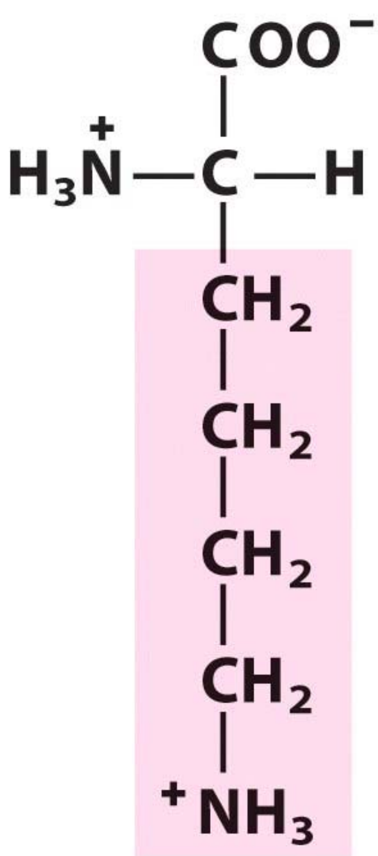


Tyrosine

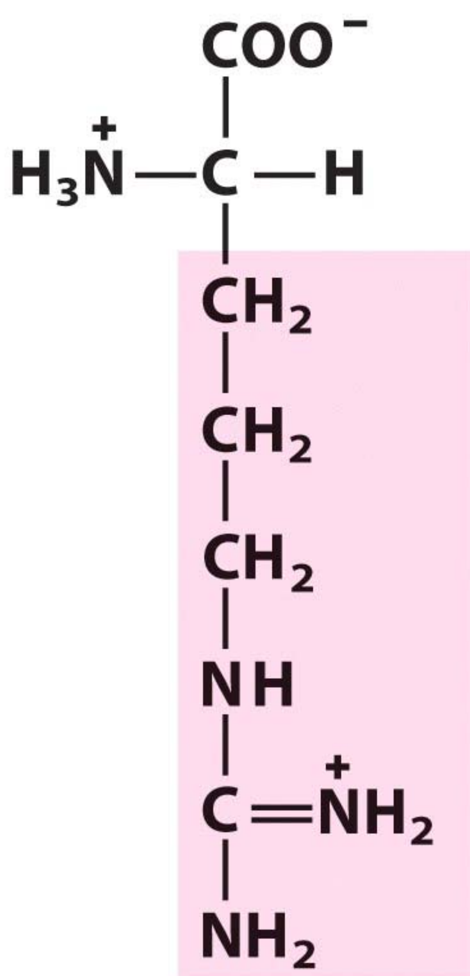


Tryptophan

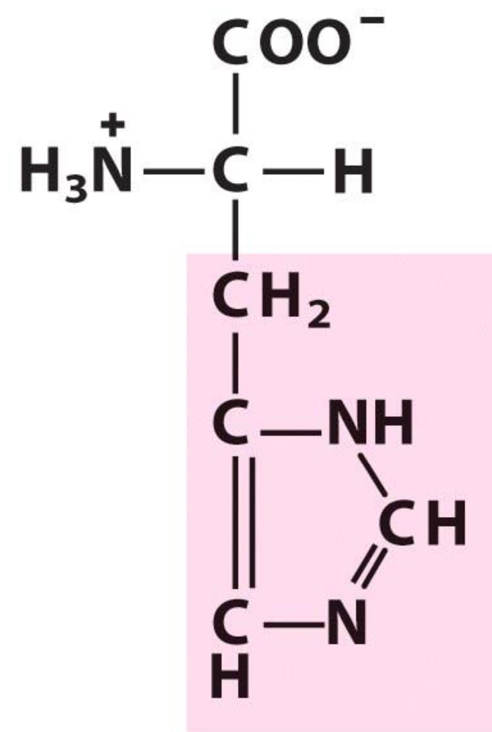
Positively charged R groups



Lysine

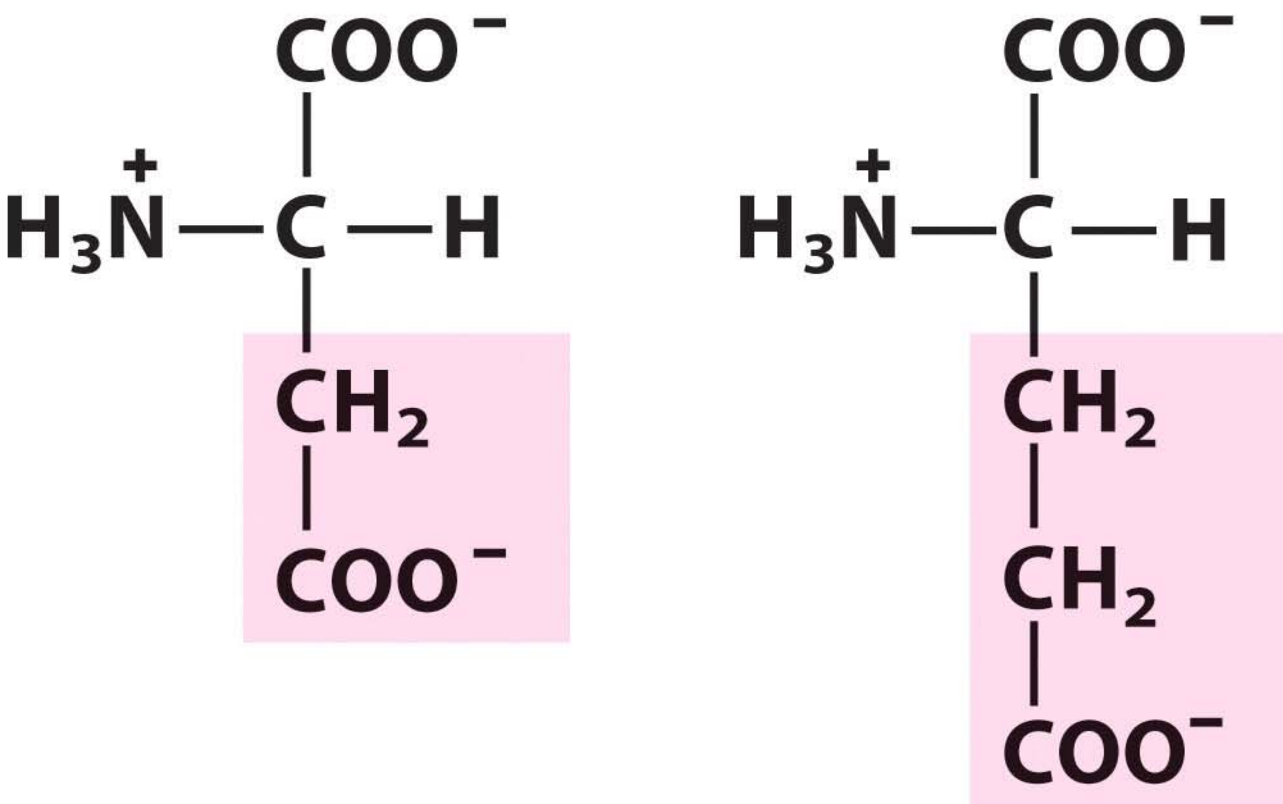


Arginine



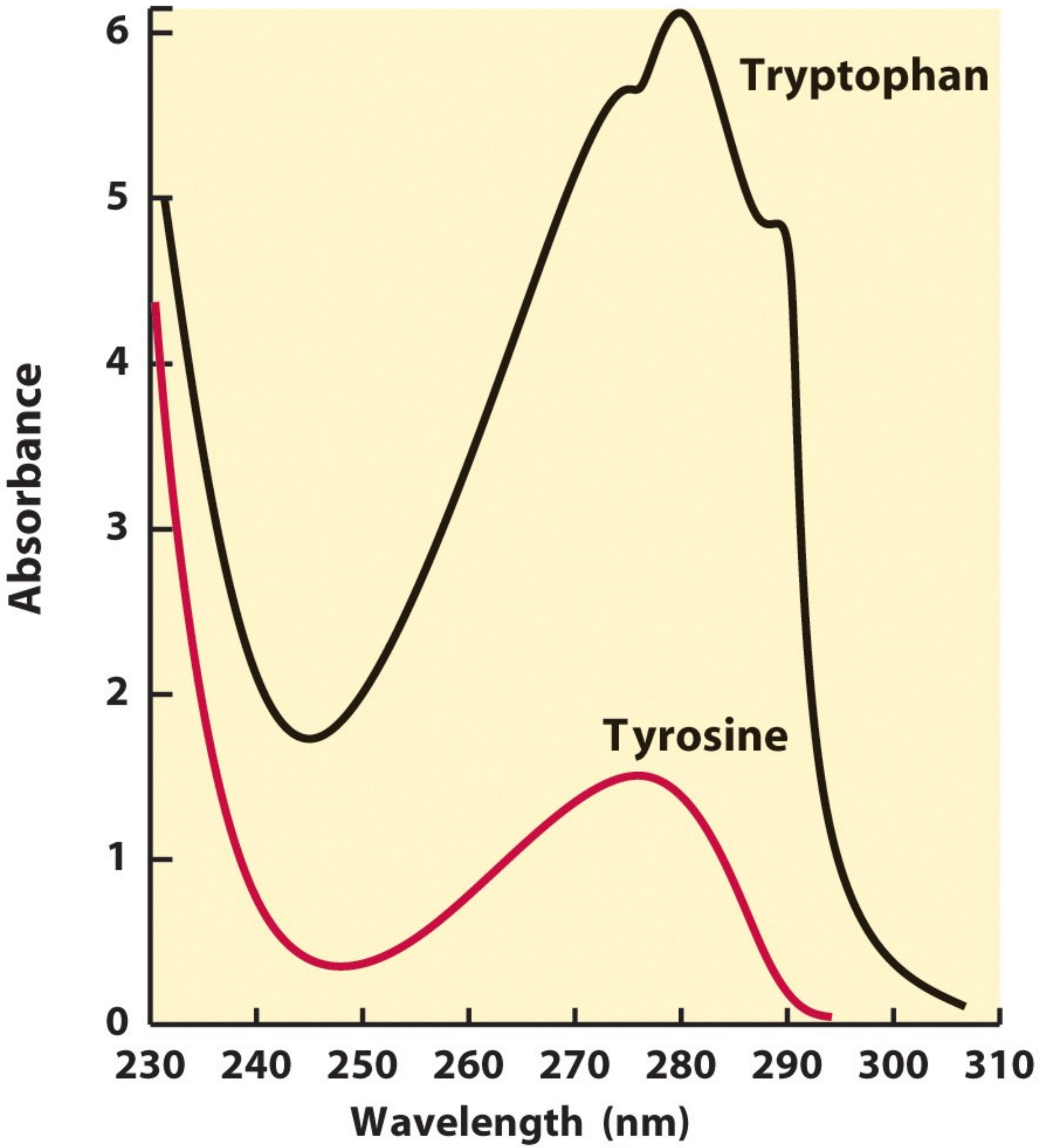
Histidine

Negatively charged R groups

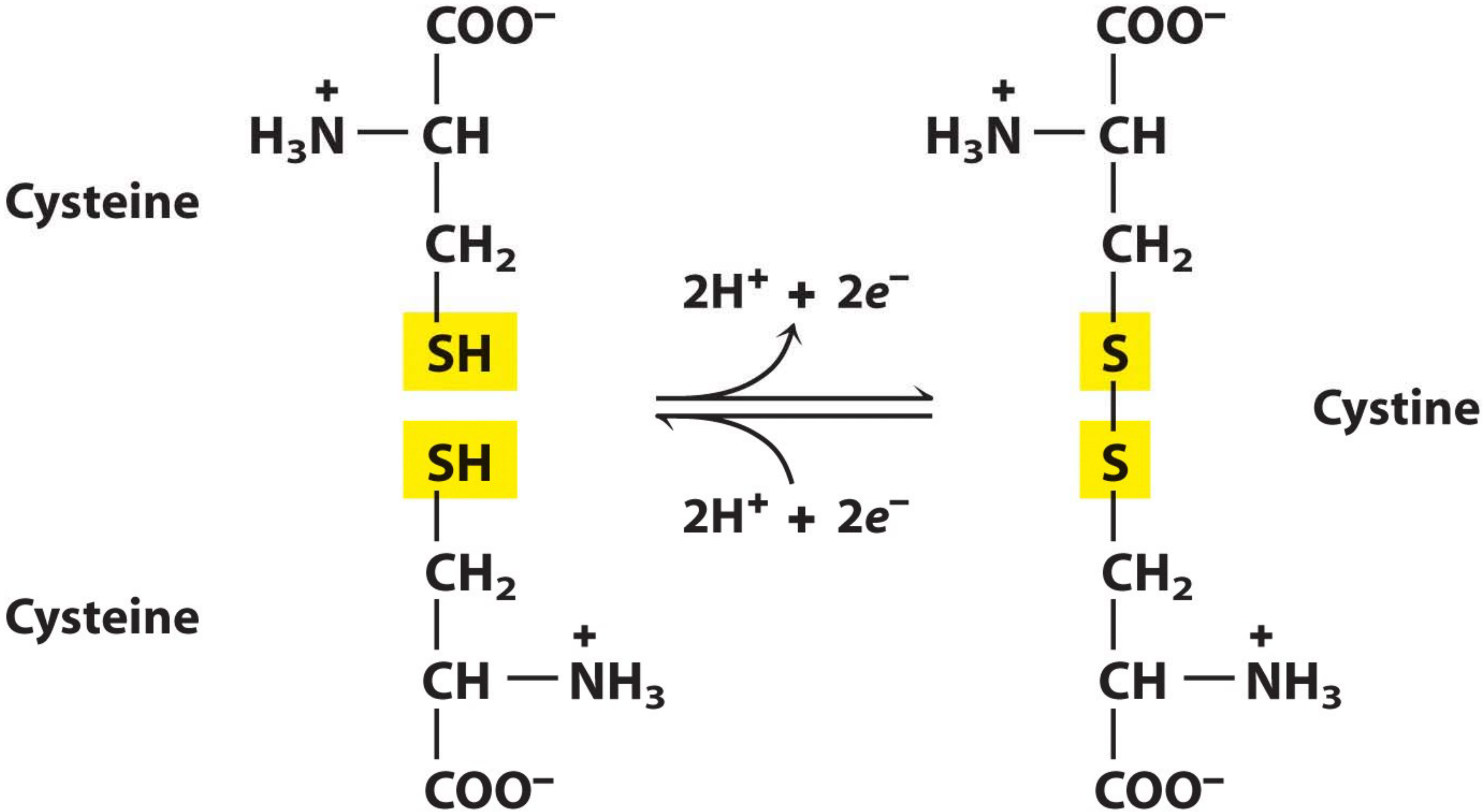


Aspartate

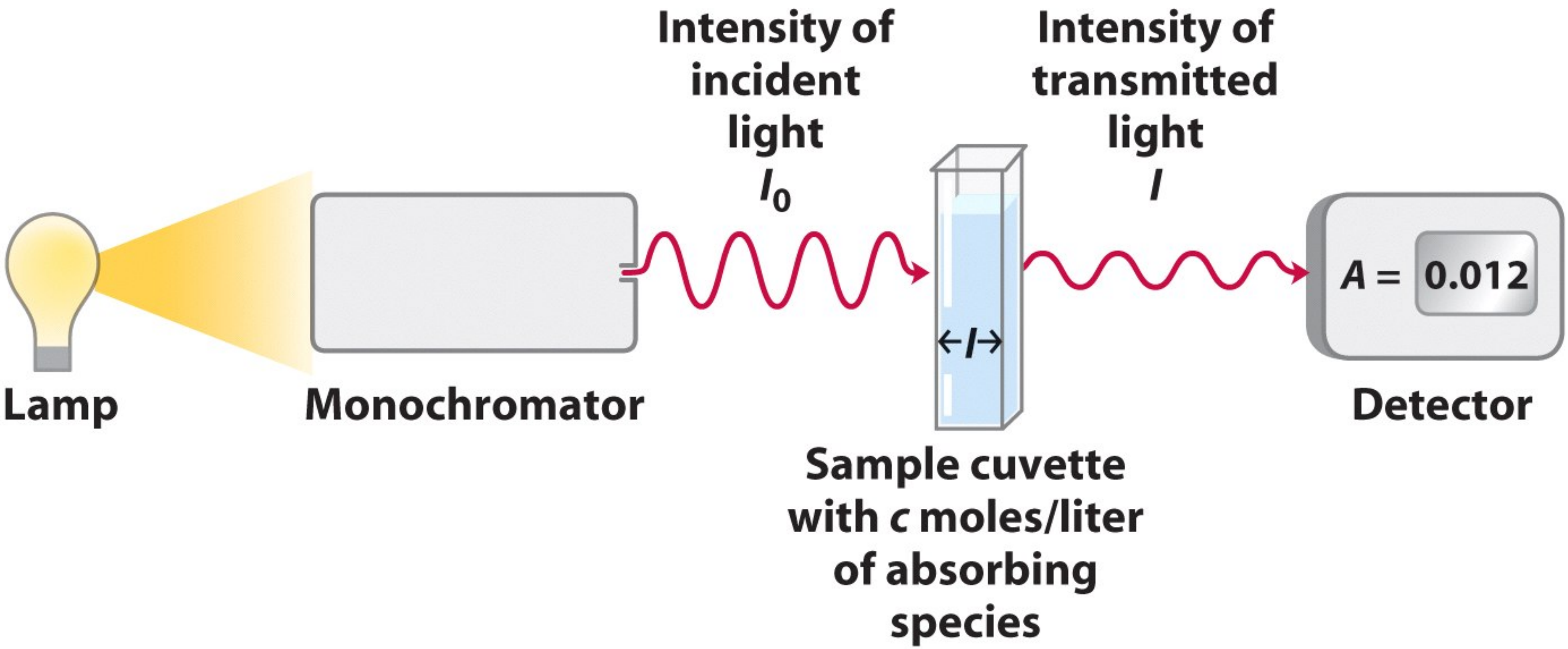
Glutamate

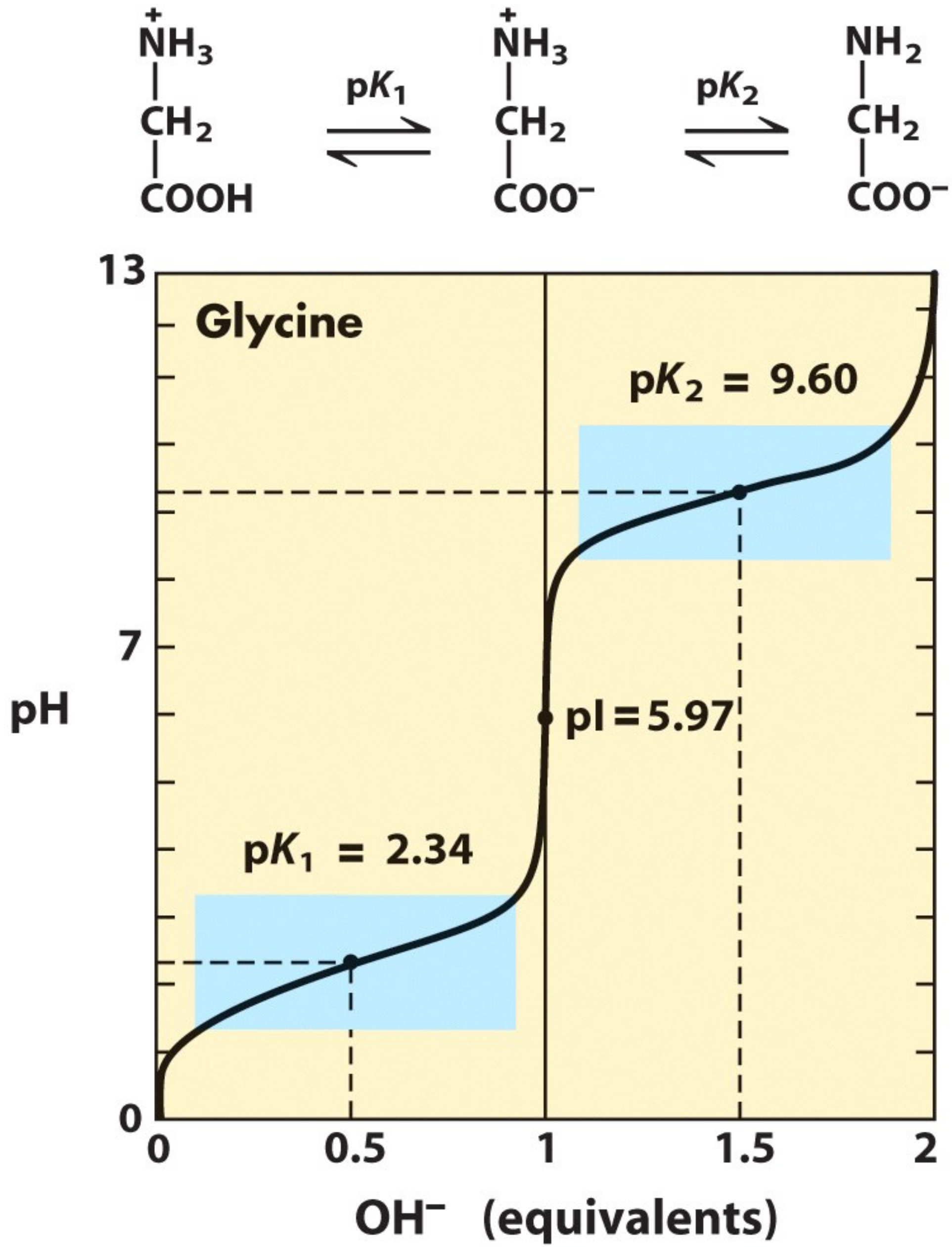


Tryptophan
Tyrosine
Absorbance
0
1
2
3
4
5
6
230
240
250
260
270
280
290
300
310
Wavelength (nm)

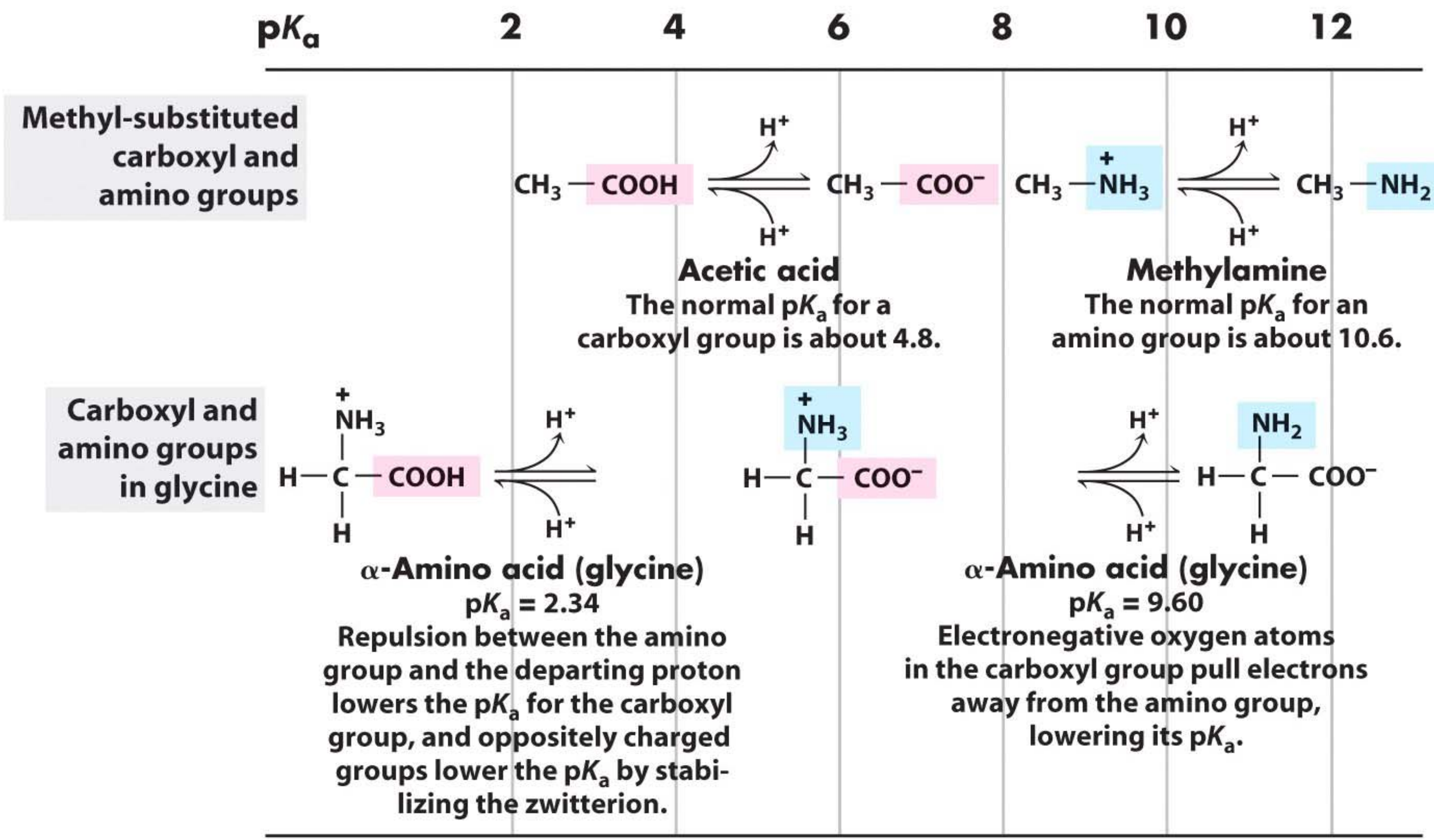


Cysteine
Cysteine
Cystine
COO⁻
H₃N⁺ — CH
CH₂
SH
SH
S
S
NH₃⁺
2H⁺ + 2e⁻
2H⁺ + 2e⁻

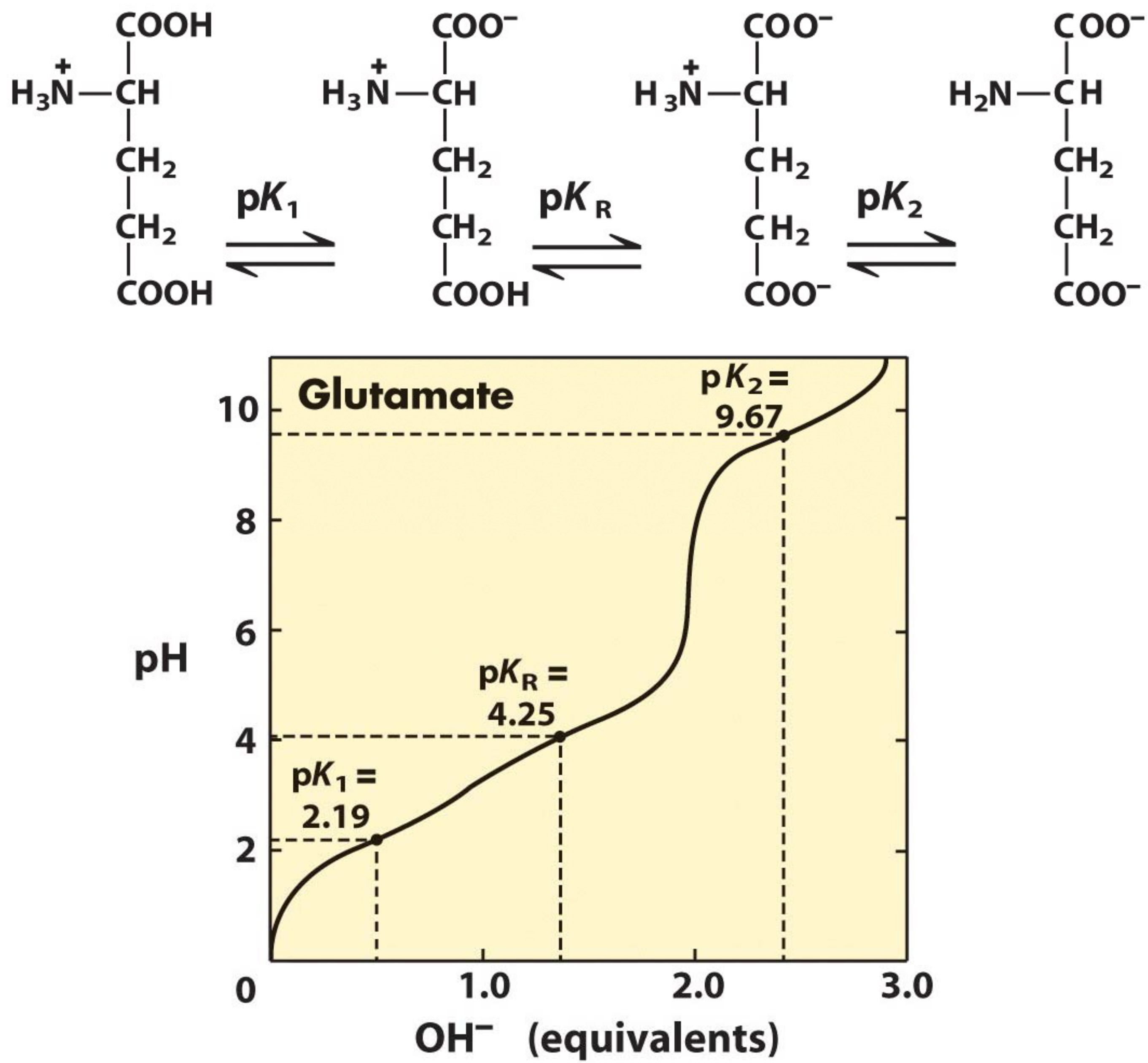

Intensity of incident light I_0
Intensity of transmitted light I
A = 0.012
←l→
Lamp
Monochromator
Detector
Sample cuvette with c moles/liter of absorbing species



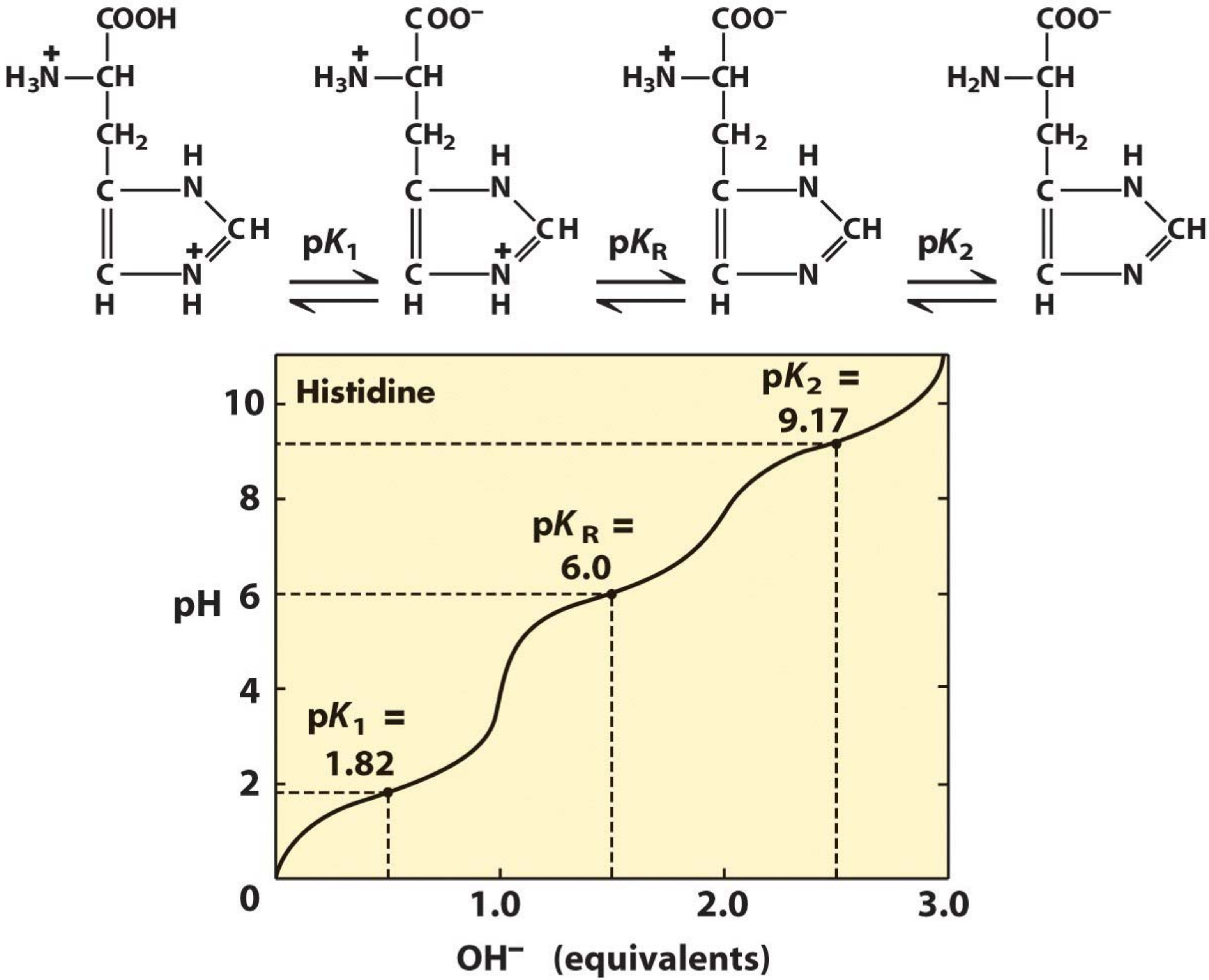
pK1
pK2
Glycine
pK2 = 9.60
pI = 5.97
pK1 = 2.34
13
7
0
pH
0
0.5
1
1.5
2
OH⁻ (equivalents)



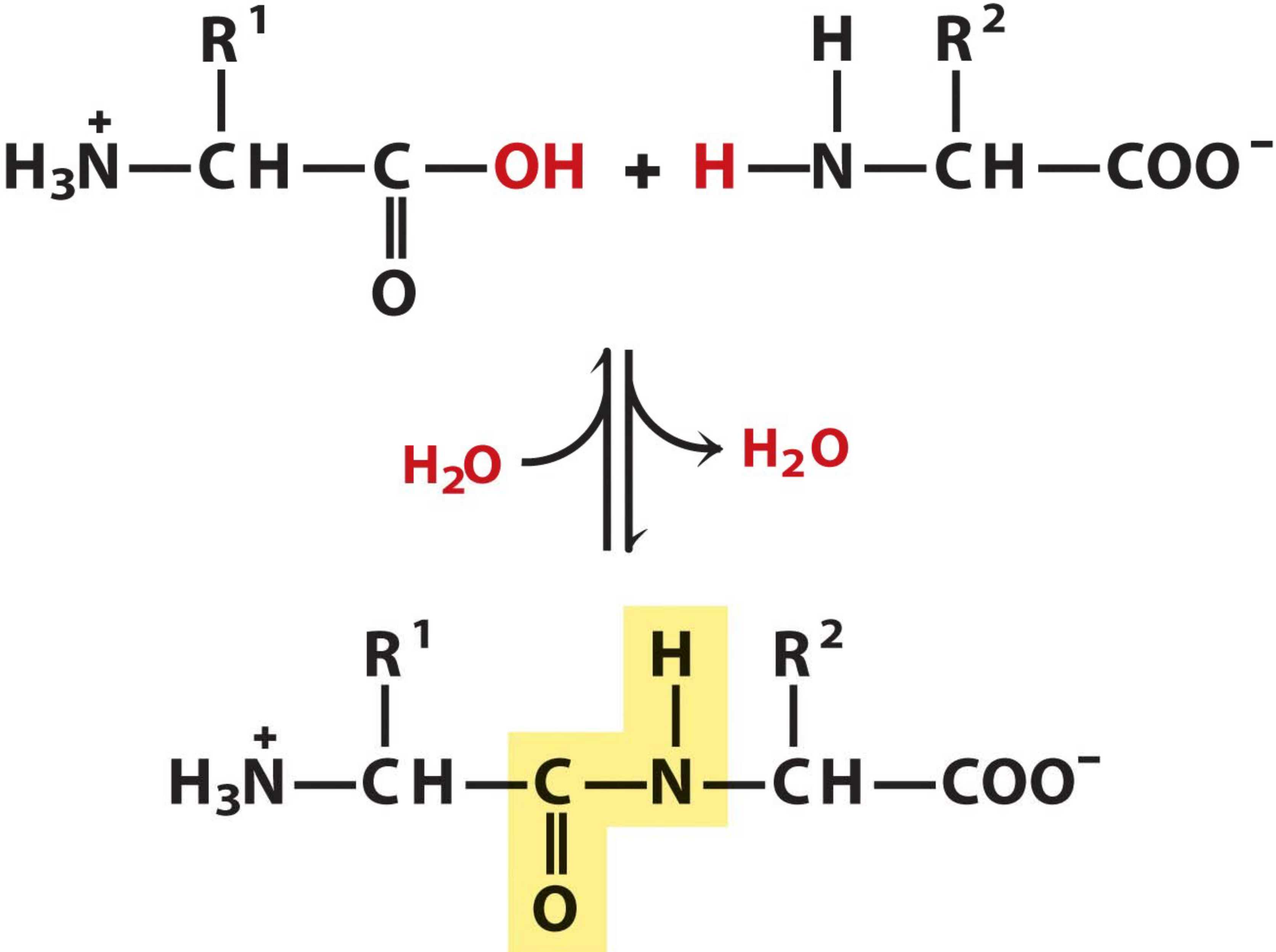
pKa
2
4
6
8
10
12
Methyl-substituted carboxyl and amino groups
CH3 — COOH ⇌ CH3 — COO⁻
H+
Acetic acid
The normal pKa for a carboxyl group is about 4.8.
CH3 — NH3+ ⇌ CH3 — NH2
Methylamine
The normal pKa for an amino group is about 10.6.
Carboxyl and amino groups in glycine
H — C — COOH
NH3+
H
H — C — COO⁻
H — C — COO⁻
NH2
α-Amino acid (glycine)
pKa = 2.34
Repulsion between the amino group and the departing proton lowers the pKa for the carboxyl group, and oppositely charged groups lower the pKa by stabilizing the zwitterion.
α-Amino acid (glycine)
pKa = 9.60
Electronegative oxygen atoms in the carboxyl group pull electrons away from the amino group, lowering its pKa.



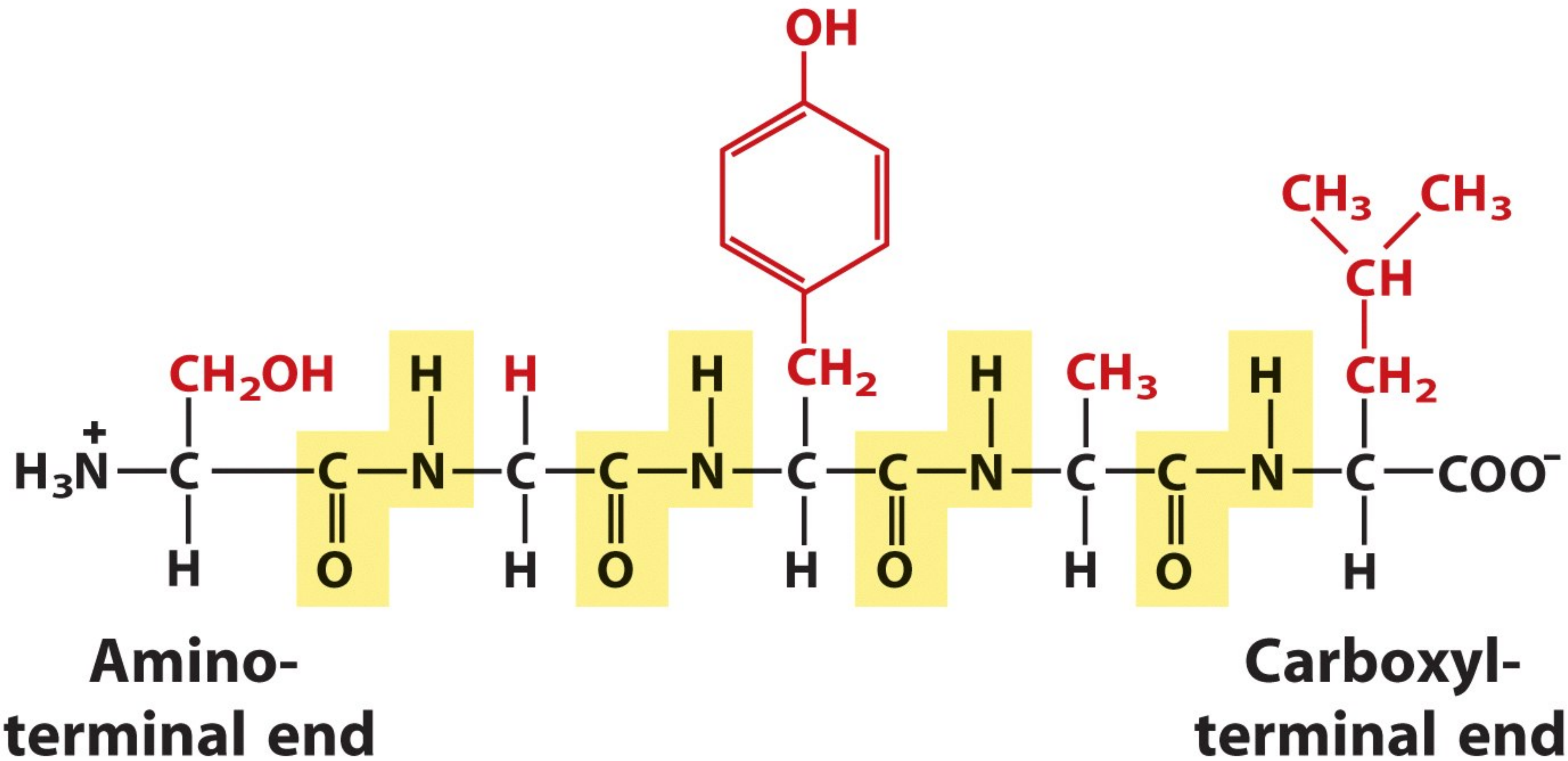
COOH
$H_3\overset{+}{N}$—CH
CH_2
CH_2
COOH
pK_1
COO^-
$H_3\overset{+}{N}$—CH
CH_2
CH_2
COOH
pK_R
COO^-
$H_3\overset{+}{N}$—CH
CH_2
CH_2
COO^-
pK_2
COO^-
H_2N—CH
CH_2
CH_2
COO^-
Glutamate
pK_2 = 9.67
pK_R = 4.25
pK_1 = 2.19
pH
0
2
4
6
8
10
1.0
2.0
3.0
OH^- (equivalents)



COOH
H3N—CH
CH2
C—N
CH
C—N
H H
pK1
COO⁻
pKR
pK2
H2N—CH
Histidine
pK2 = 9.17
pKR = 6.0
pK1 = 1.82
pH
10
8
6
4
2
0
1.0
2.0
3.0
OH⁻ (equivalents)



R1
H
R2
H3N+ — CH — C — OH + H — N — CH — COO−
O
H2O
H2O
R1
H
R2
H3N+ — CH — C — N — CH — COO−
O


Amino-terminal end
Carboxyl-terminal end

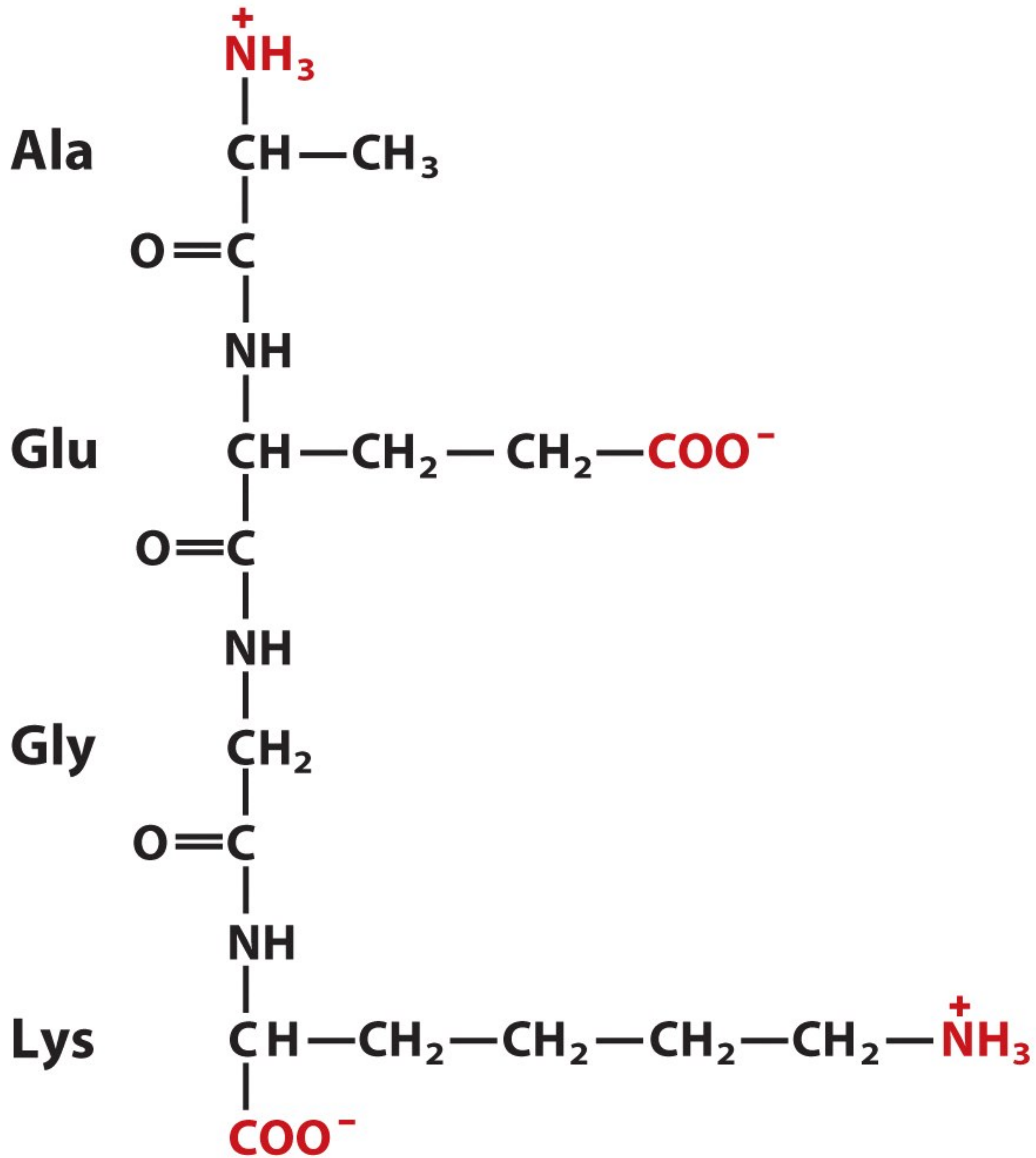

Ala
Glu
Gly
Lys
NH3
CH—CH3
O=C
NH
CH—CH2—CH2—COO⁻
O=C
NH
CH2
O=C
NH
CH—CH2—CH2—CH2—CH2—NH3
COO⁻

TABLE 3–2 Molecular Data on Some Proteins

	Molecular weight	*Number of residues*	*Number of polypeptide chains*
Cytochrome *c* (human)	13,000	104	1
Ribonuclease A (bovine pancreas)	13,700	124	1
Lysozyme (chicken egg white)	13,930	129	1
Myoglobin (equine heart)	16,890	153	1
Chymotrypsin (bovine pancreas)	21,600	241	3
Chymotrypsinogen (bovine)	22,000	245	1
Hemoglobin (human)	64,500	574	4
Serum albumin (human)	68,500	609	1
Hexokinase (yeast)	102,000	972	2
RNA polymerase (*E. coli*)	450,000	4,158	5
Apolipoprotein B (human)	513,000	4,536	1
Glutamine synthetase (*E. coli*)	619,000	5,628	12
Titin (human)	2,993,000	26,926	1

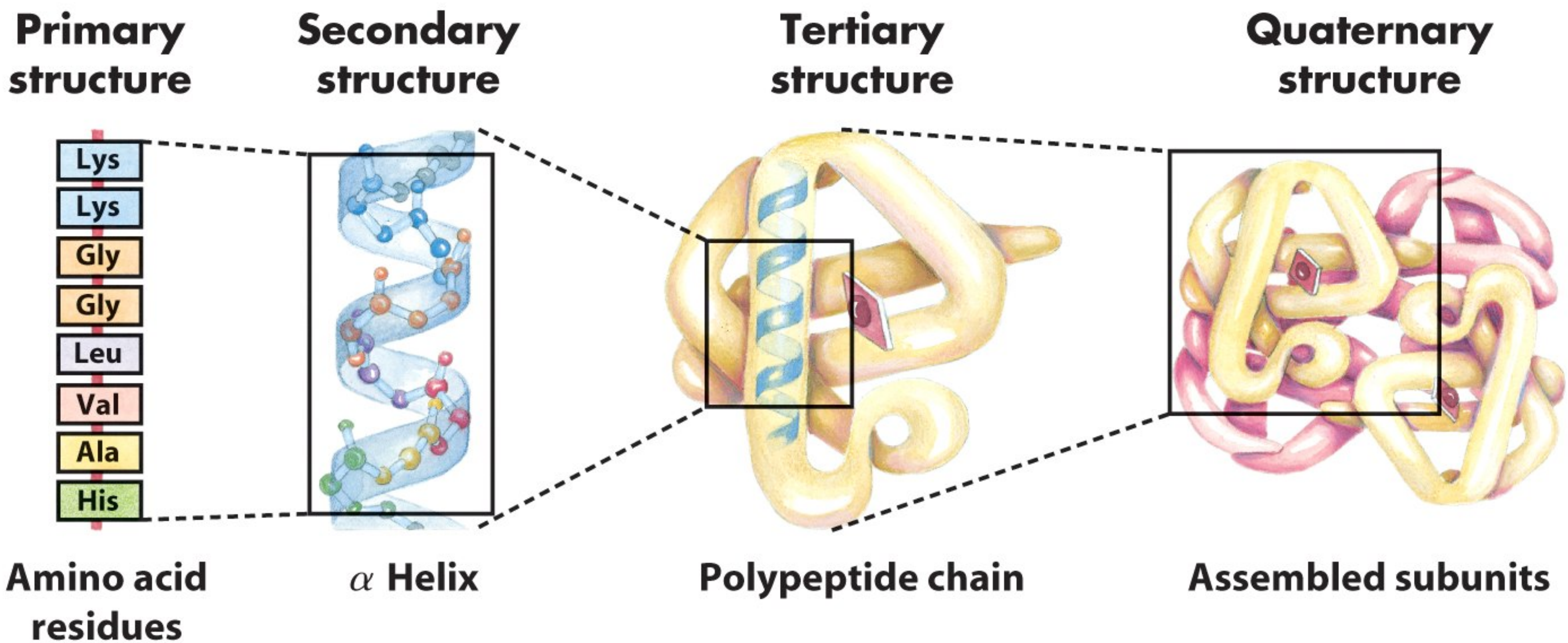
TABLE 3–3 Amino Acid Composition of Two Proteins

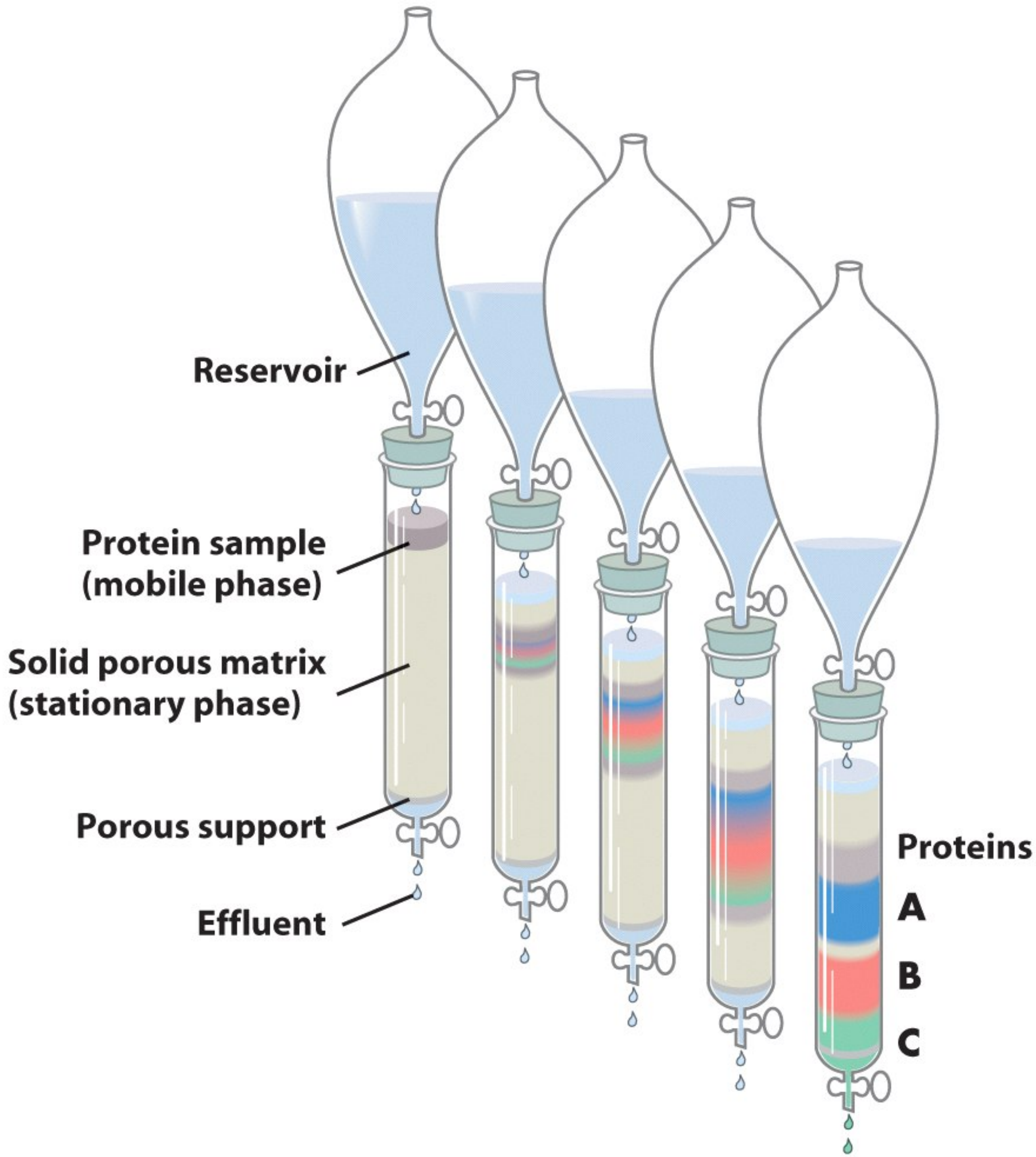
	*Number of residues per molecule of protein**	
Amino acid	*Bovine cytochrome* c	*Bovine chymotrypsinogen*
Ala	6	22
Arg	2	4
Asn	5	15
Asp	3	8
Cys	2	10
Gln	3	10
Glu	9	5
Gly	14	23
His	3	2
Ile	6	10
Leu	6	19
Lys	18	14
Met	2	2
Phe	4	6
Pro	4	9
Ser	1	28
Thr	8	23
Trp	1	8
Tyr	4	4
Val	3	23
Total	104	245

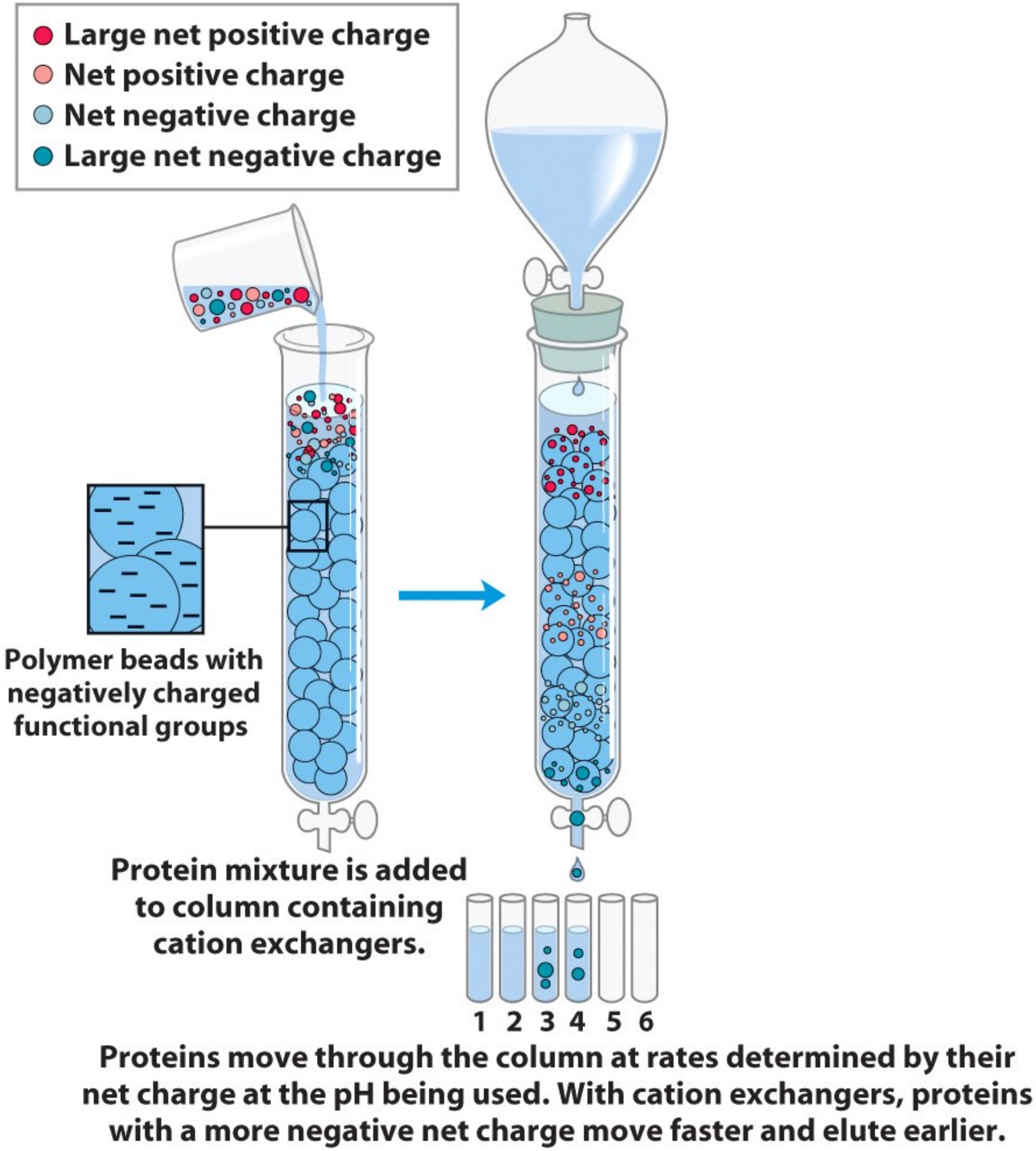
*In some common analyses, such as acid hydrolysis, Asp and Asn are not readily distinguished from each other and are together designated Asx (or B). Similarly, when Glu and Gln cannot be distinguished, they are together designated Glx (or Z). In addition, Trp is destroyed. Additional procedures must be employed to obtain an accurate assessment of complete amino acid content.

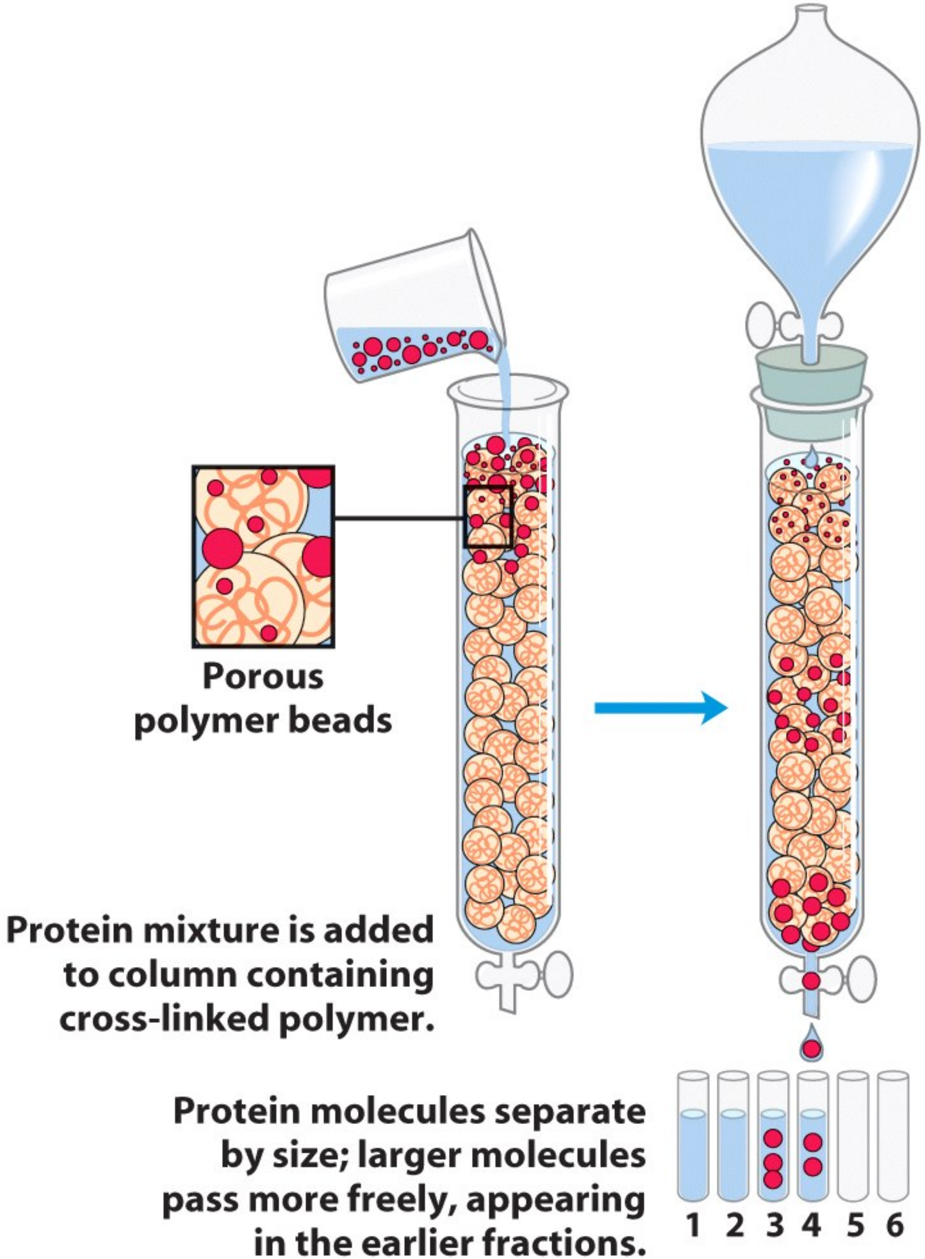
TABLE 3–4 Conjugated Proteins

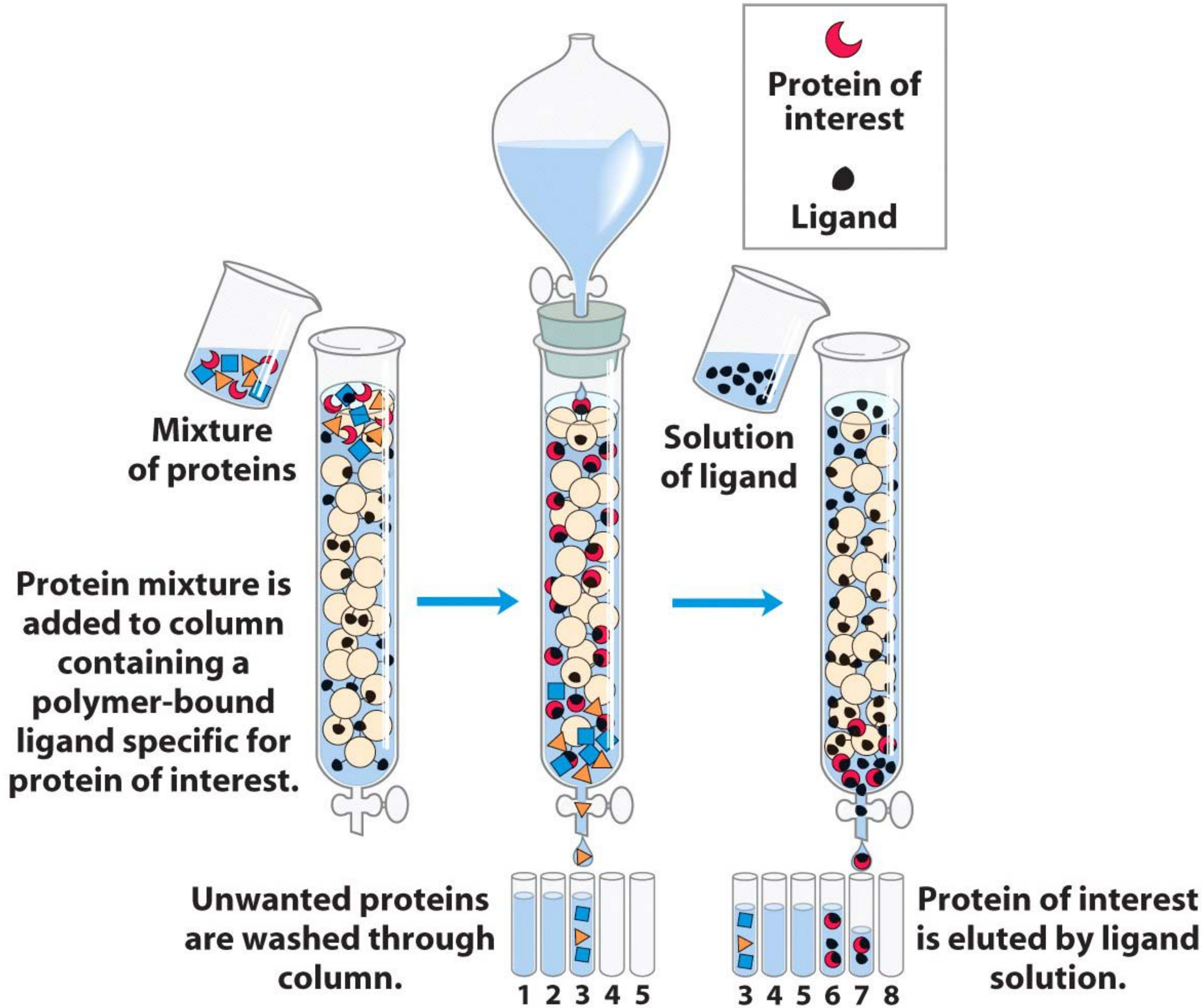
Class	*Prosthetic group*	*Example*
Lipoproteins	Lipids	β_1-Lipoprotein of blood
Glycoproteins	Carbohydrates	Immunoglobulin G
Phosphoproteins	Phosphate groups	Casein of milk
Hemoproteins	Heme (iron porphyrin)	Hemoglobin
Flavoproteins	Flavin nucleotides	Succinate dehydrogenase
Metalloproteins	Iron	Ferritin
	Zinc	Alcohol dehydrogenase
	Calcium	Calmodulin
	Molybdenum	Dinitrogenase
	Copper	Plastocyanin


Primary structure
Secondary structure
Tertiary structure
Quaternary structure
Lys
Lys
Gly
Gly
Leu
Val
Ala
His
Amino acid residues
α Helix
Polypeptide chain
Assembled subunits


Reservoir
Protein sample
(mobile phase)
Solid porous matrix
(stationary phase)
Porous support
Effluent
Proteins
A
B
C


Large net positive charge
Net positive charge
Net negative charge
Large net negative charge
Polymer beads with negatively charged functional groups
Protein mixture is added to column containing cation exchangers.
1 2 3 4 5 6
Proteins move through the column at rates determined by their net charge at the pH being used. With cation exchangers, proteins with a more negative net charge move faster and elute earlier.


Porous
polymer beads
Protein mixture is added
to column containing
cross-linked polymer.
Protein molecules separate
by size; larger molecules
pass more freely, appearing
in the earlier fractions.
1 2 3 4 5 6

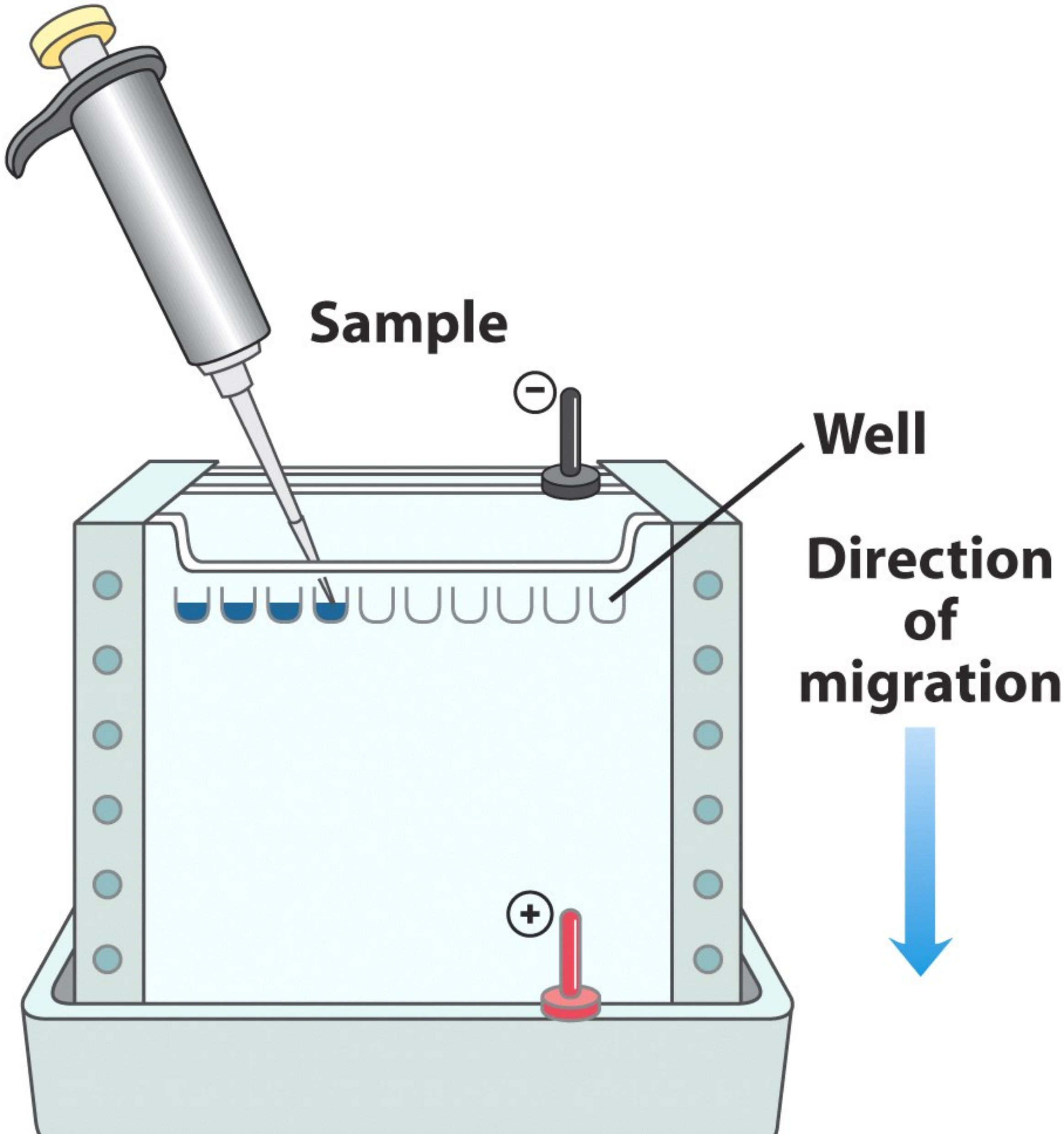


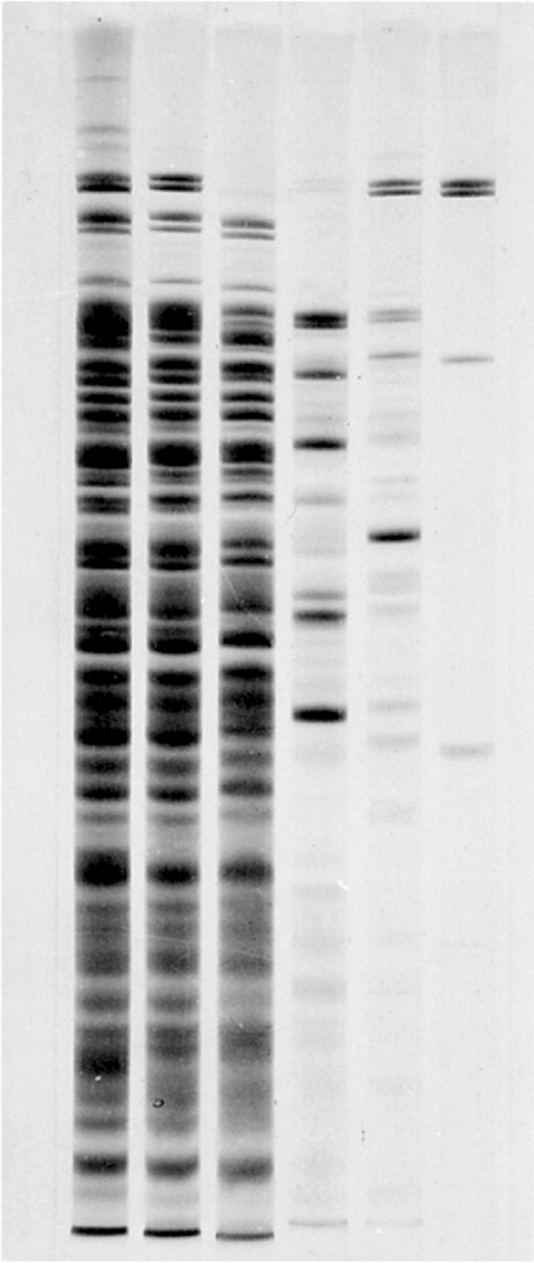
Protein of interest
Ligand
Mixture of proteins
Protein mixture is added to column containing a polymer-bound ligand specific for protein of interest.
Solution of ligand
Unwanted proteins are washed through column.
1 2 3 4 5
3 4 5 6 7 8
Protein of interest is eluted by ligand solution.

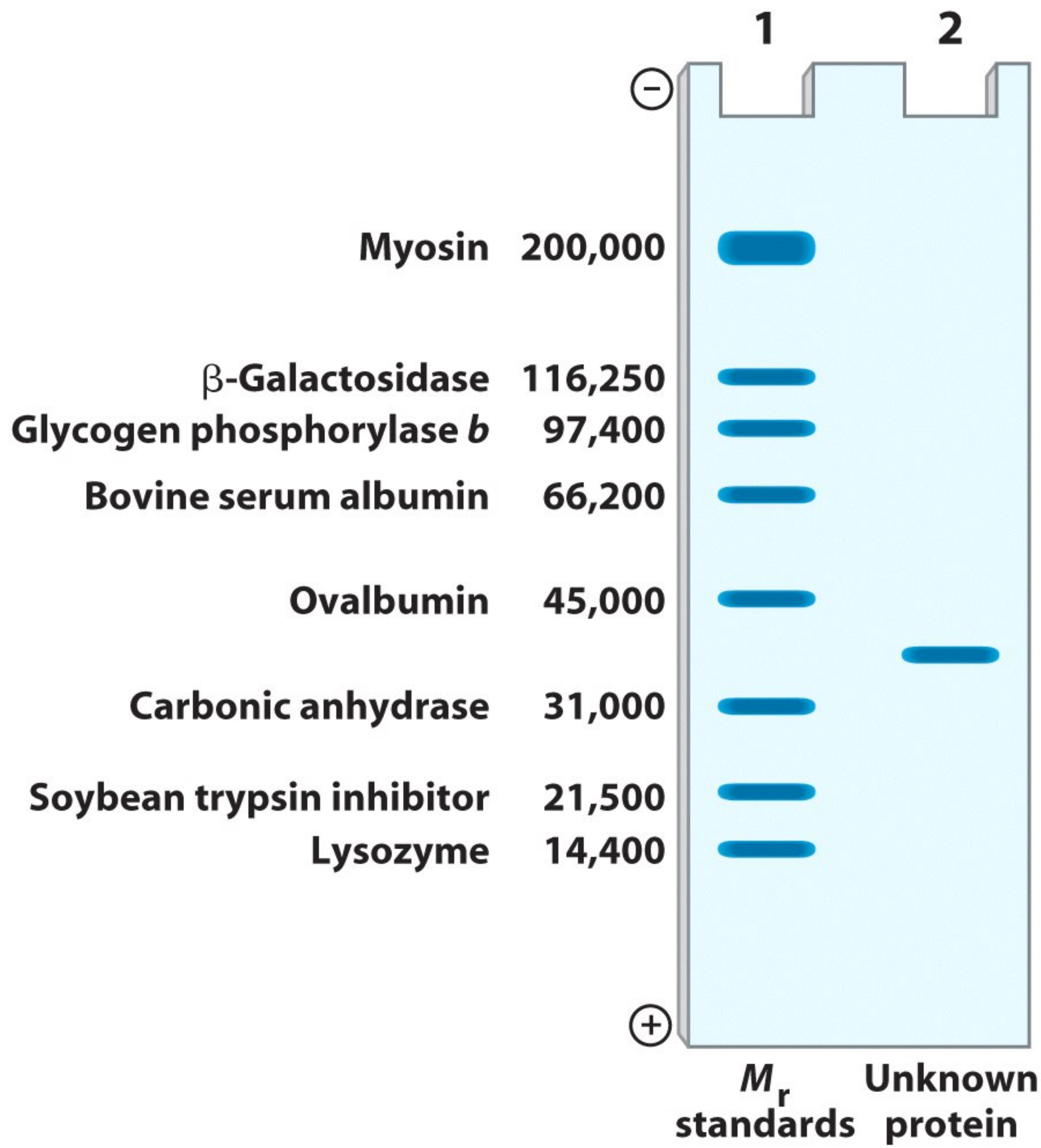
TABLE 3–5 **A Purification Table for a Hypothetical Enzyme**

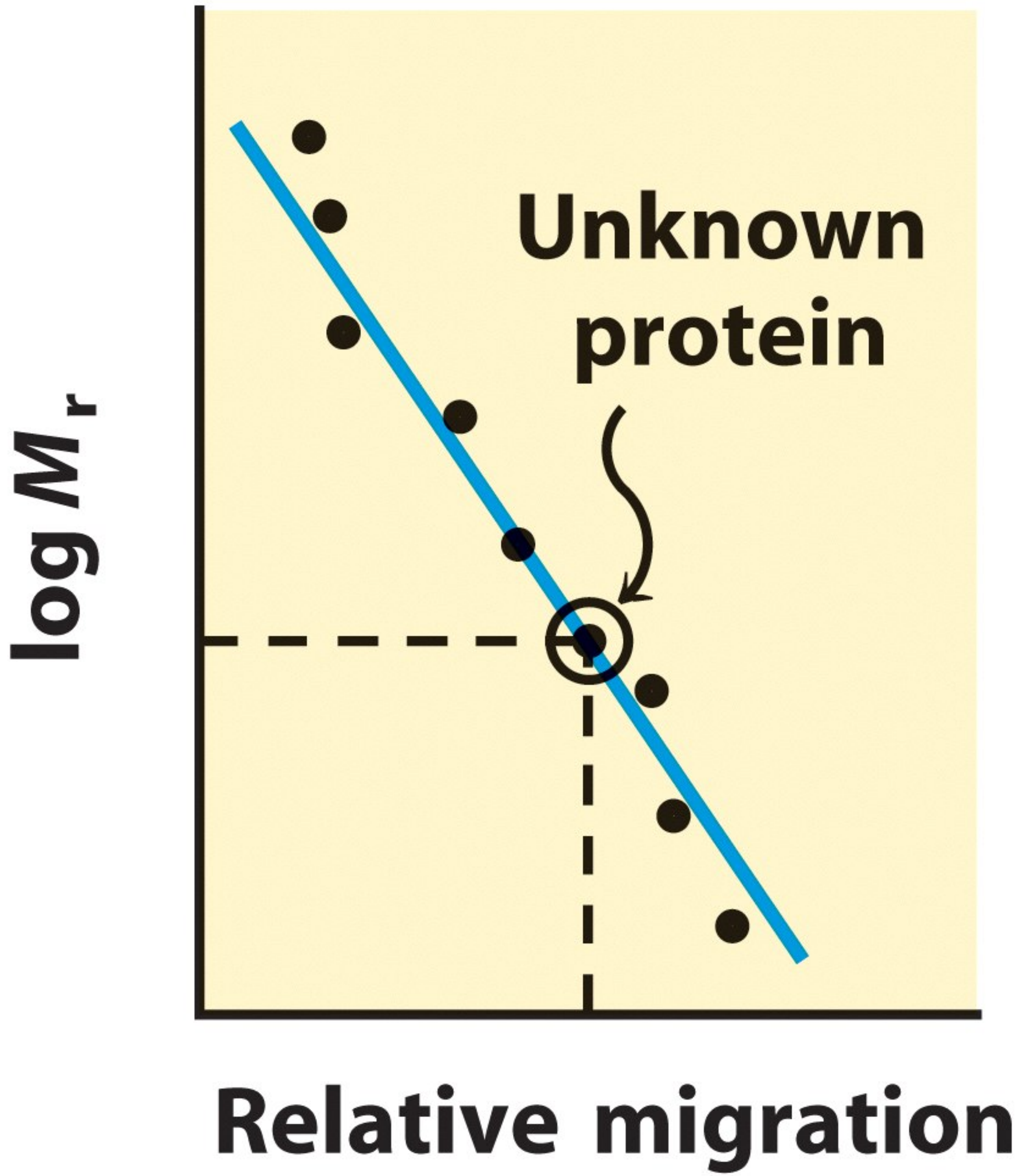
Procedure or step	*Fraction volume (ml)*	*Total protein (mg)*	*Activity (units)*	*Specific activity (units/mg)*
1. Crude cellular extract	1,400	10,000	100,000	10
2. Precipitation with ammonium sulfate	280	3,000	96,000	32
3. Ion-exchange chromatography	90	400	80,000	200
4. Size-exclusion chromatography	80	100	60,000	600
5. Affinity chromatography	6	3	45,000	15,000

Note: All data represent the status of the sample *after* the designated procedure has been carried out. Activity and specific activity are defined on page 94.


Sample
Well
Direction of migration




1
2
Myosin 200,000
β-Galactosidase 116,250
Glycogen phosphorylase *b* 97,400
Bovine serum albumin 66,200
Ovalbumin 45,000
Carbonic anhydrase 31,000
Soybean trypsin inhibitor 21,500
Lysozyme 14,400
M_r standards
Unknown protein



Unknown protein
log M_r
Relative migration

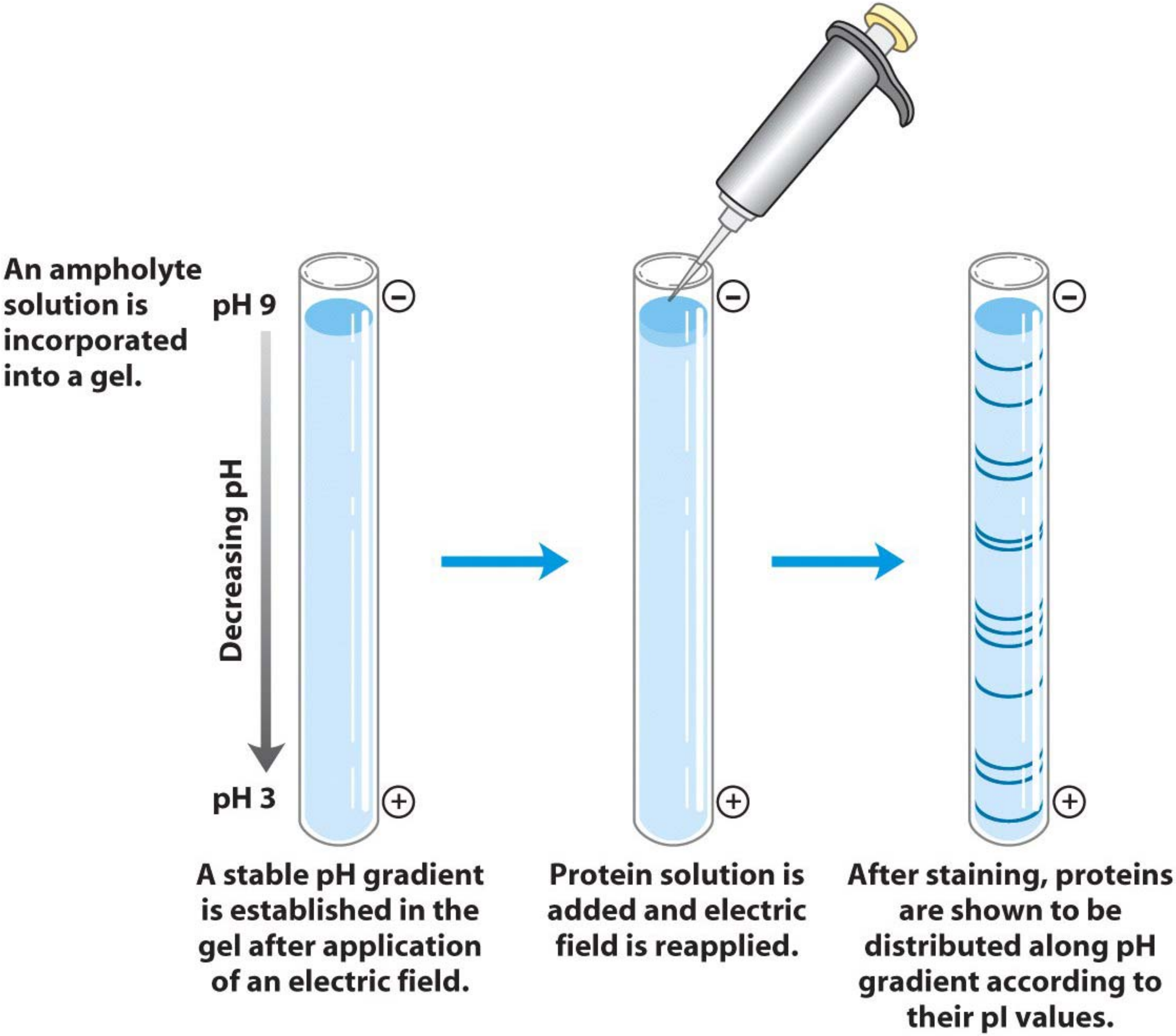
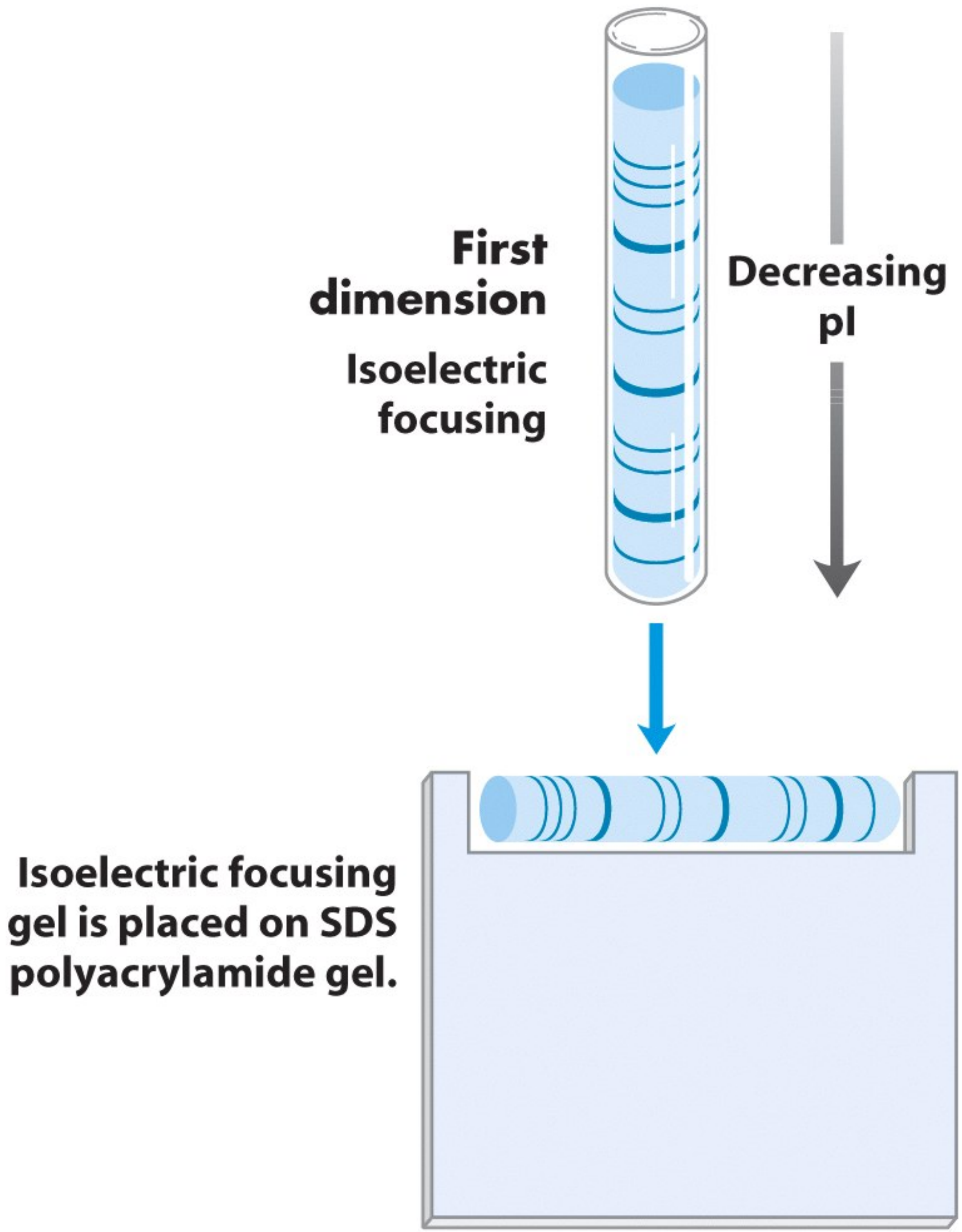
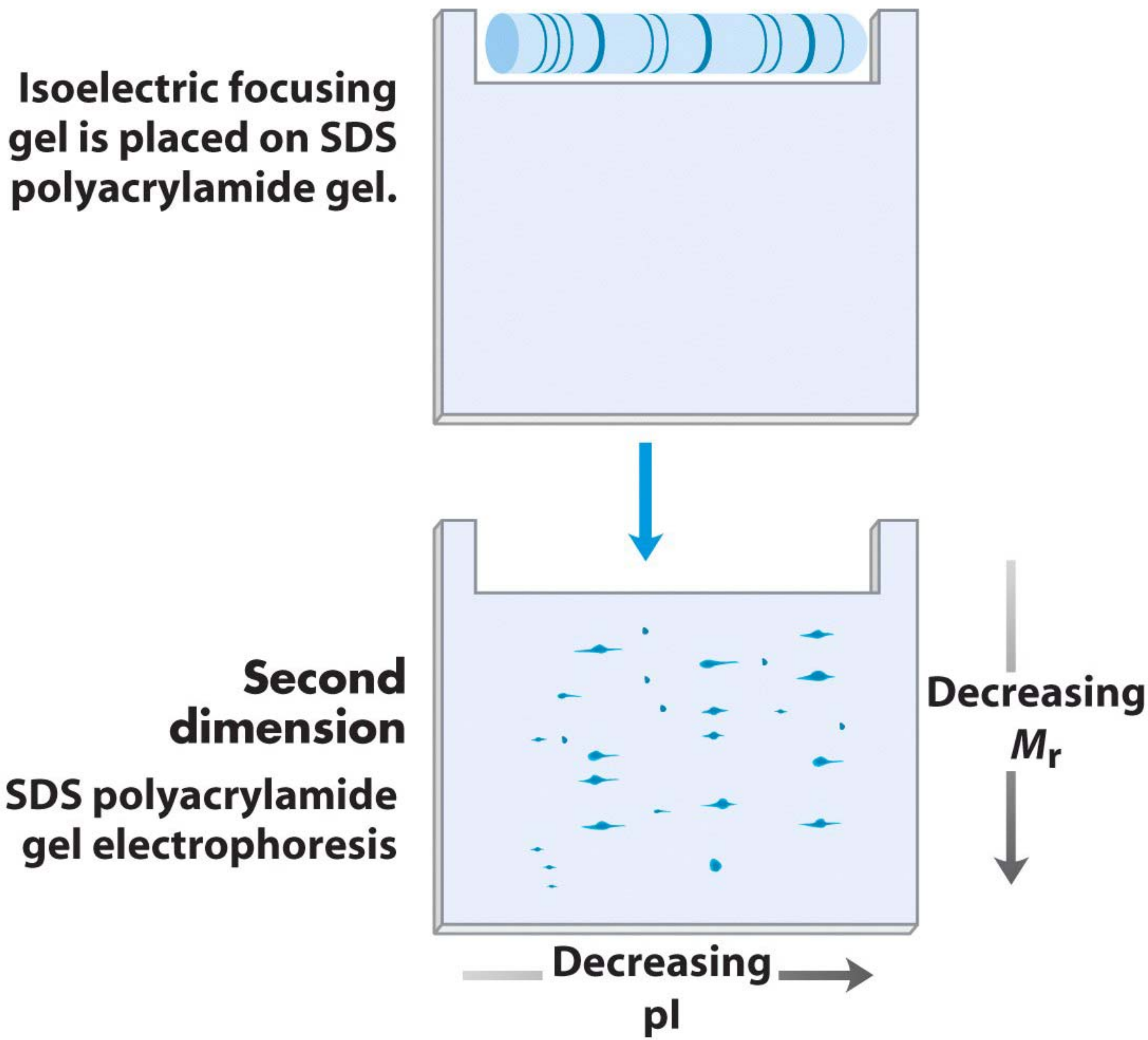
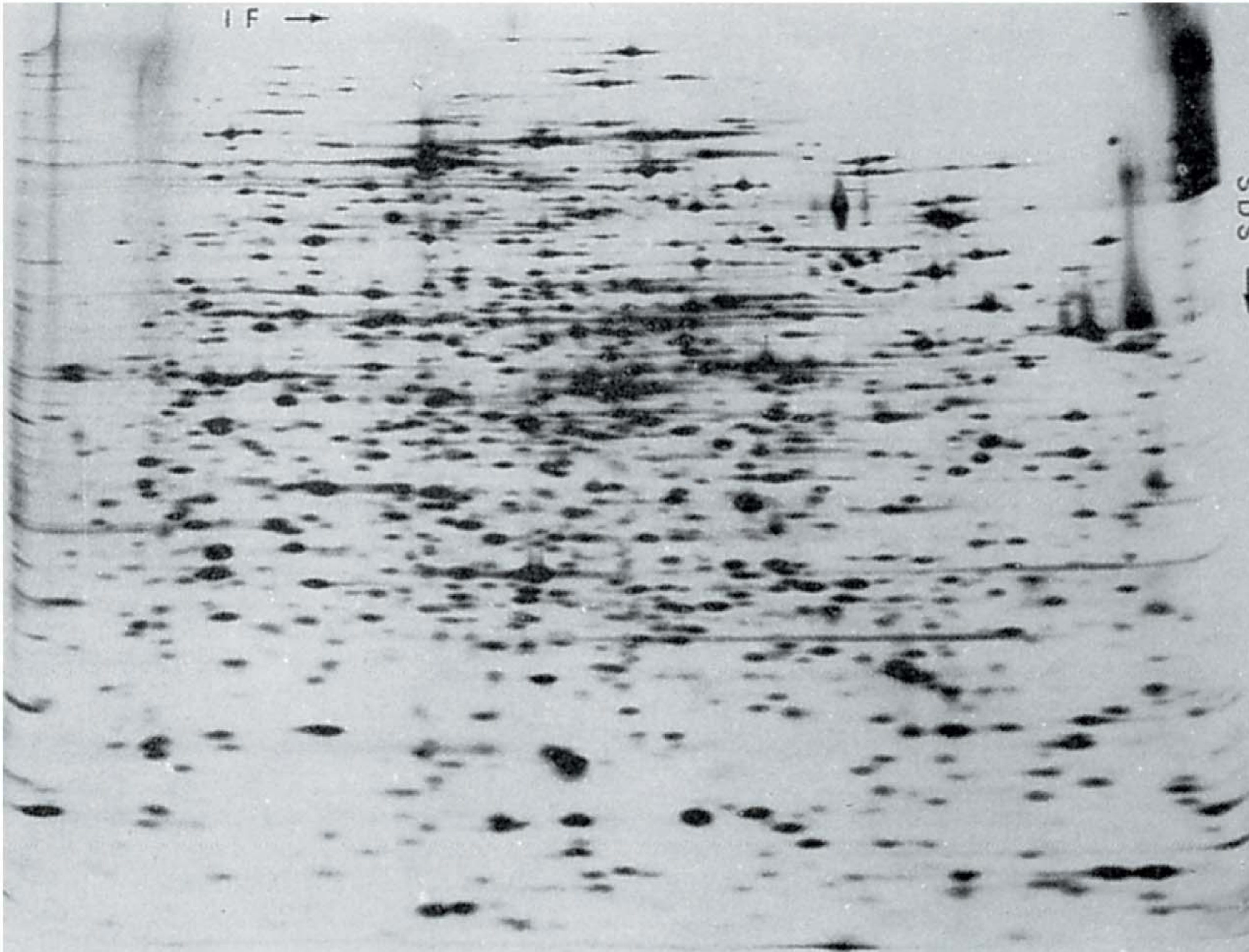

An ampholyte solution is incorporated into a gel.
pH 9
Decreasing pH
pH 3
A stable pH gradient is established in the gel after application of an electric field.
Protein solution is added and electric field is reapplied.
After staining, proteins are shown to be distributed along pH gradient according to their pI values.

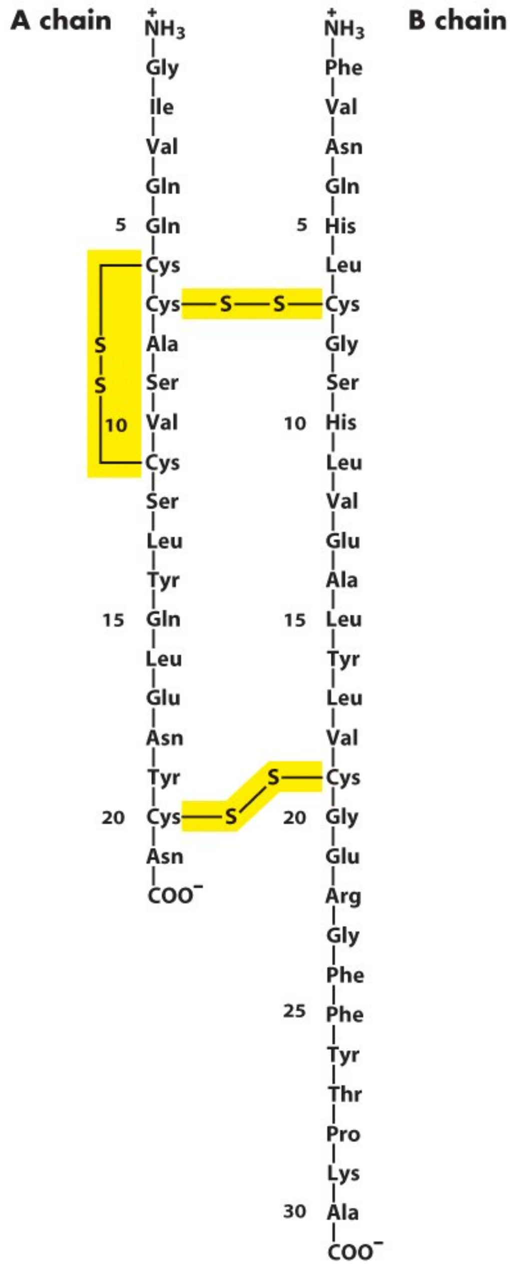
TABLE 3–6 The Isoelectric Points of Some Proteins

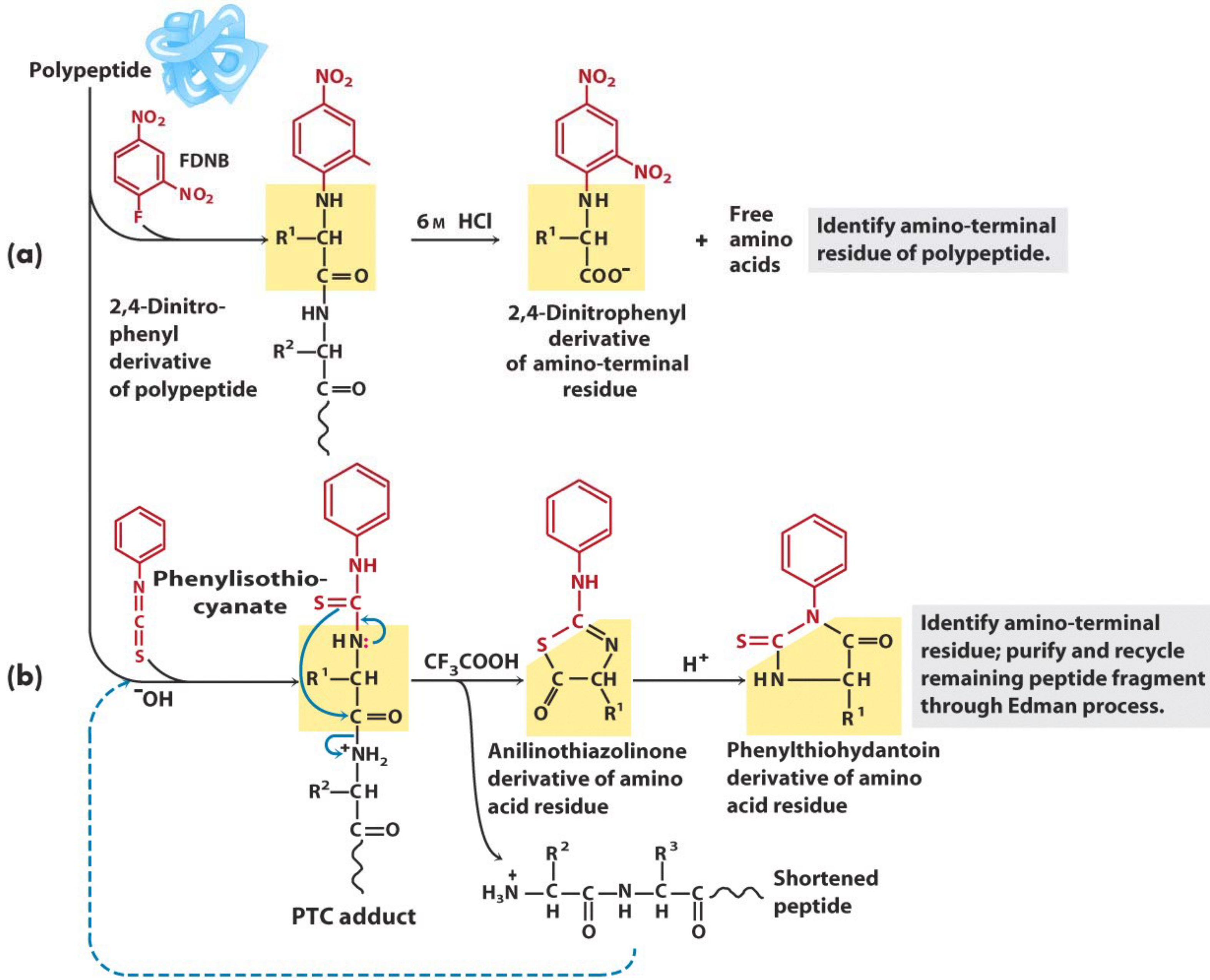
Protein	*pI*
Pepsin	<1.0
Egg albumin	4.6
Serum albumin	4.9
Urease	5.0
β-Lactoglobulin	5.2
Hemoglobin	6.8
Myoglobin	7.0
Chymotrypsinogen	9.5
Cytochrome *c*	10.7
Lysozyme	11.0


First
dimension
Isoelectric
focusing
Decreasing
pI
Isoelectric focusing
gel is placed on SDS
polyacrylamide gel.


Isoelectric focusing gel is placed on SDS polyacrylamide gel.
Second dimension
SDS polyacrylamide gel electrophoresis
Decreasing M_r
Decreasing pI


IF →
SDS →


A chain
B chain
NH3+
Gly
Ile
Val
Gln
5
Gln
Cys
Cys
S
S
Ala
Ser
10
Val
Cys
Ser
Leu
Tyr
15
Gln
Leu
Glu
Asn
Tyr
20
Cys
Asn
COO−
Phe
Val
Asn
Gln
His
Leu
Cys
Gly
Ser
His
Leu
Val
Glu
Ala
Leu
Tyr
Leu
Val
Cys
Gly
Glu
Arg
Gly
Phe
25
Phe
Tyr
Thr
Pro
Lys
30
Ala
S
S
S
S


Polypeptide
FDNB
(a)
2,4-Dinitrophenyl derivative of polypeptide
6 M HCl
2,4-Dinitrophenyl derivative of amino-terminal residue
+ Free amino acids
Identify amino-terminal residue of polypeptide.
(b)
Phenylisothiocyanate
⁻OH
CF3COOH
H+
PTC adduct
Anilinothiazolinone derivative of amino acid residue
Phenylthiohydantoin derivative of amino acid residue
Identify amino-terminal residue; purify and recycle remaining peptide fragment through Edman process.
Shortened peptide

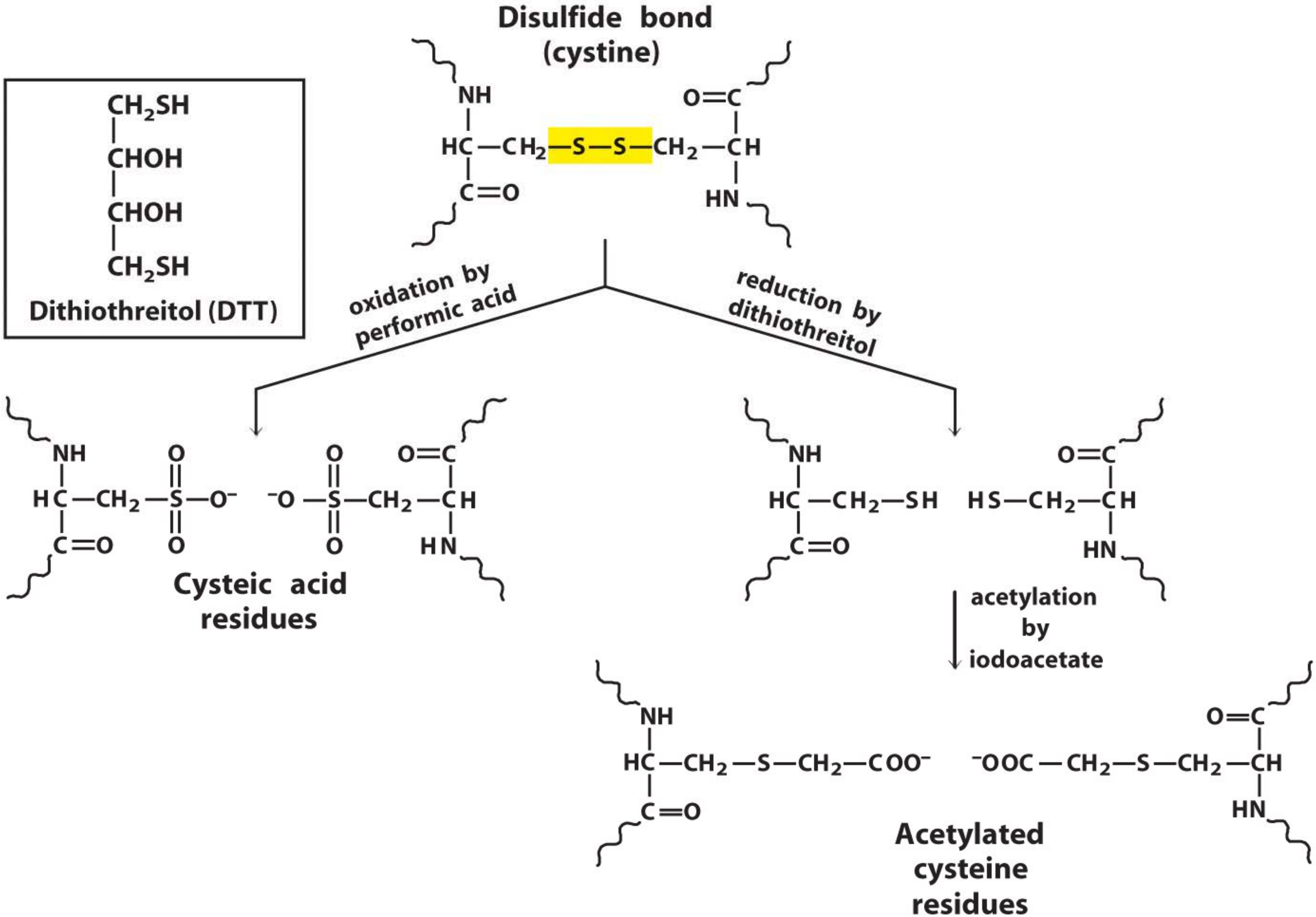
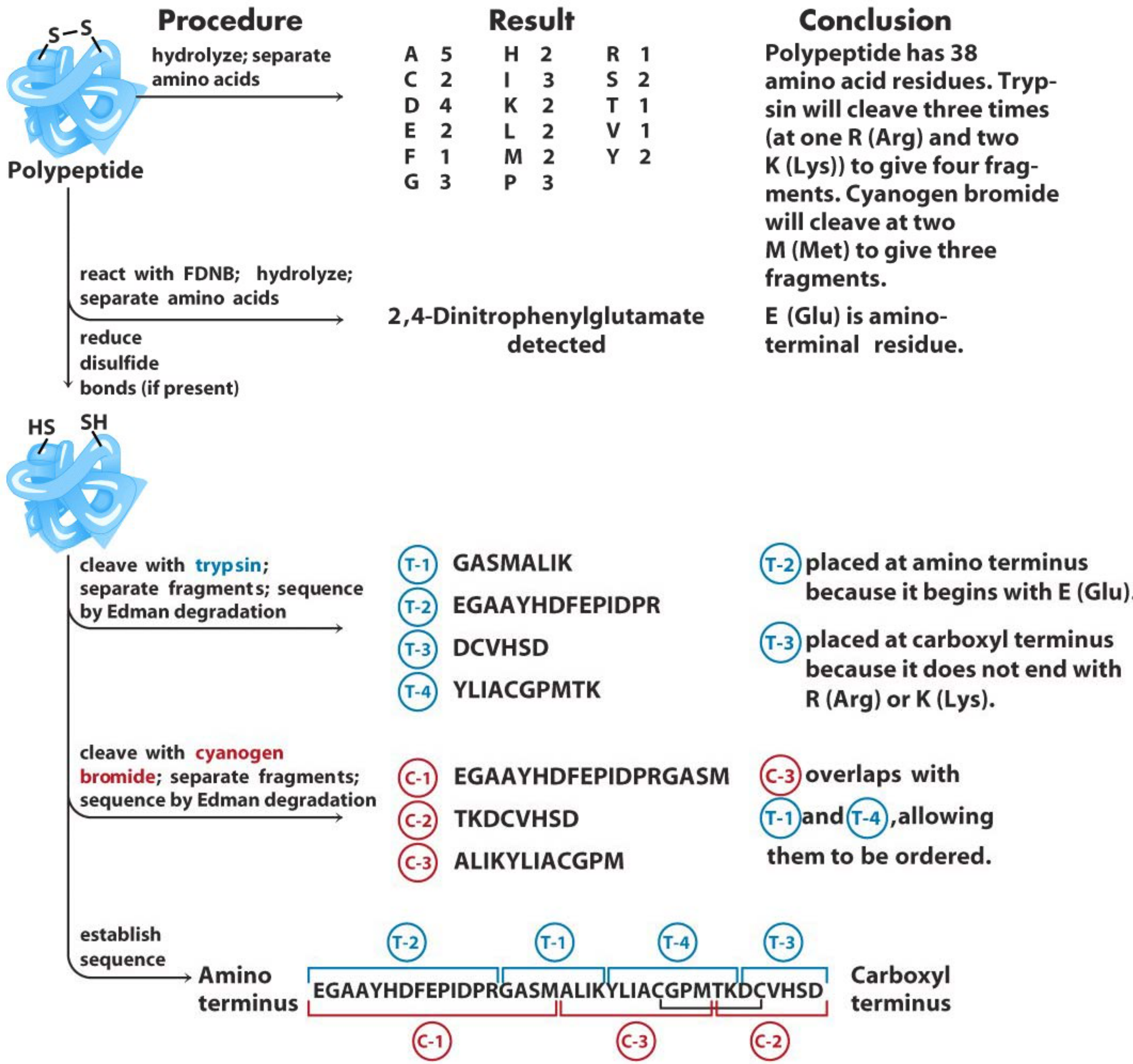

Disulfide bond
(cystine)
Dithiothreitol (DTT)
oxidation by
performic acid
reduction by
dithiothreitol
Cysteic acid
residues
acetylation
by
iodoacetate
Acetylated
cysteine
residues

TABLE 3–7 The Specificity of Some Common Methods for Fragmenting Polypeptide Chains

*Reagent (biological source)**	*Cleavage points*†
Trypsin (bovine pancreas)	Lys, Arg (C)
Submaxillarus protease (mouse submaxillary gland)	Arg (C)
Chymotrypsin (bovine pancreas)	Phe, Trp, Tyr (C)
Staphylococcus aureus V8 protease (bacterium *S. aureus*)	Asp, Glu (C)
Asp-*N*-protease (bacterium *Pseudomonas fragi*)	Asp, Glu (N)
Pepsin (porcine stomach)	Phe, Trp, Tyr (N)
Endoproteinase Lys C (bacterium *Lysobacter enzymogenes*)	Lys (C)
Cyanogen bromide	Met (C)

*All reagents except cyanogen bromide are proteases. All are available from commercial sources.

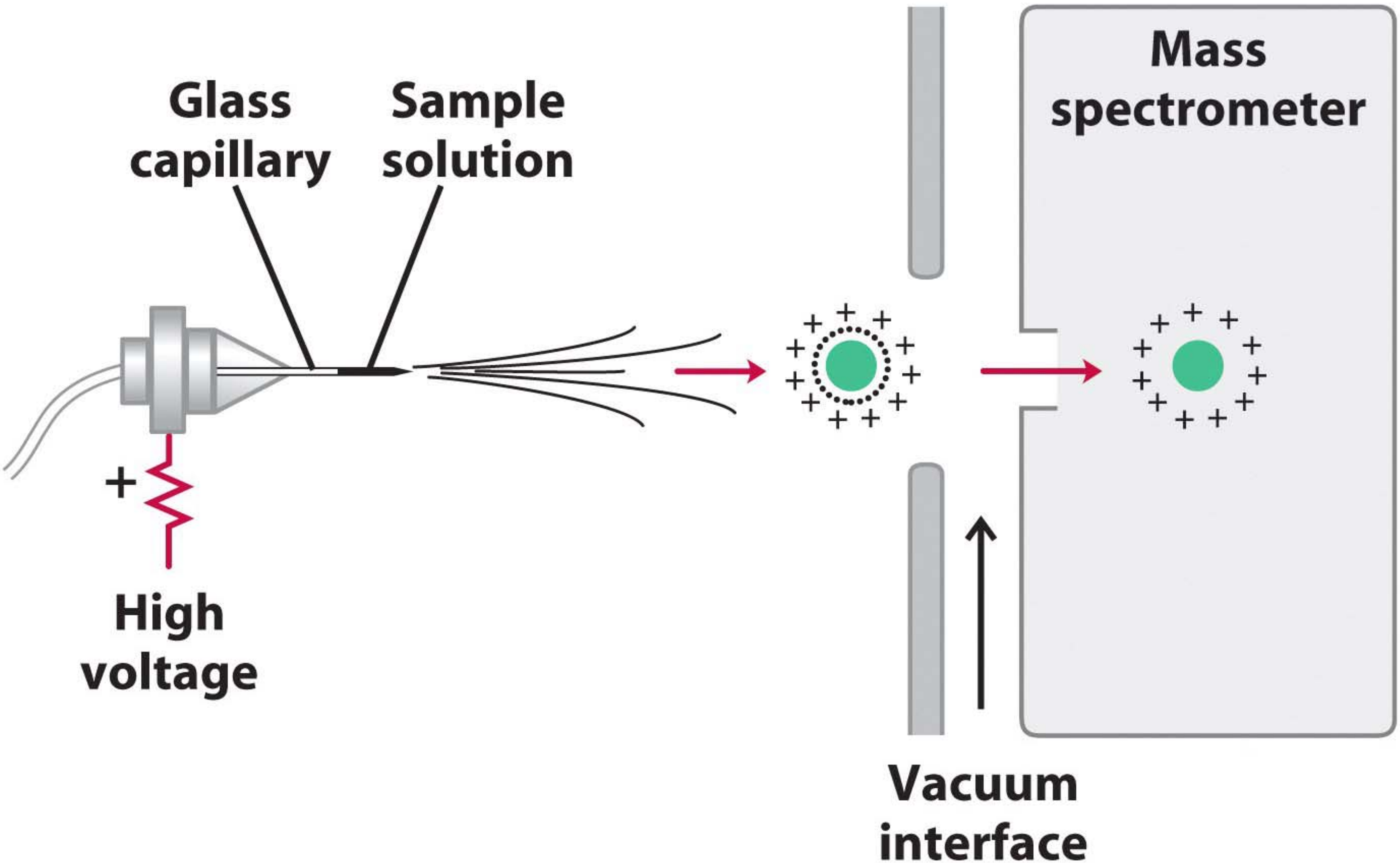
†Residues furnishing the primary recognition point for the protease or reagent; peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.

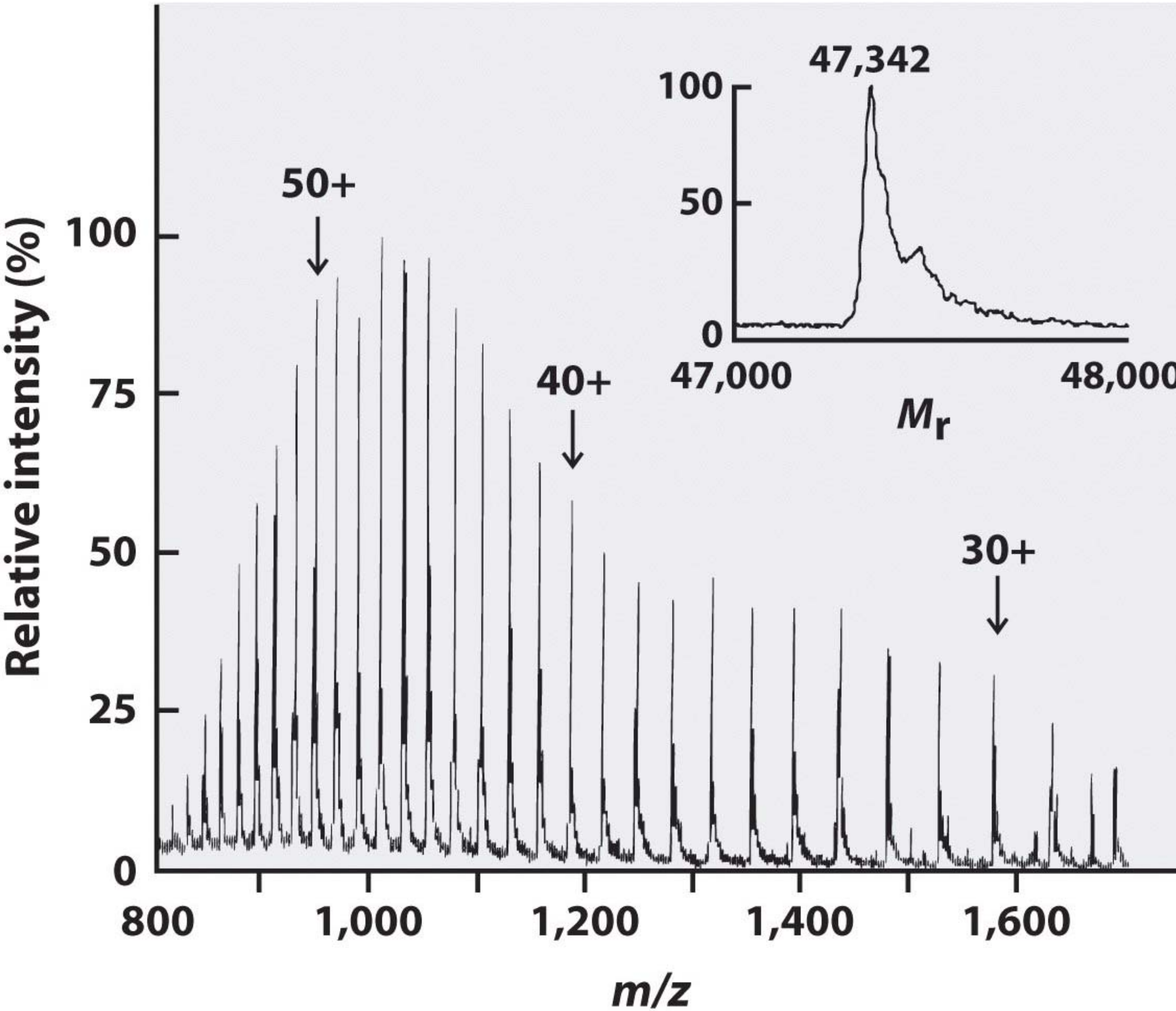


Procedure
Result
Conclusion
S–S
Polypeptide
hydrolyze; separate amino acids
A 5
C 2
D 4
E 2
F 1
G 3
H 2
I 3
K 2
L 2
M 2
P 3
R 1
S 2
T 1
V 1
Y 2
Polypeptide has 38 amino acid residues. Trypsin will cleave three times (at one R (Arg) and two K (Lys)) to give four fragments. Cyanogen bromide will cleave at two M (Met) to give three fragments.
react with FDNB; hydrolyze; separate amino acids
2,4-Dinitrophenylglutamate detected
E (Glu) is amino-terminal residue.
reduce disulfide bonds (if present)
HS
SH
cleave with trypsin; separate fragments; sequence by Edman degradation
T-1 GASMALIK
T-2 EGAAYHDFEPIDPR
T-3 DCVHSD
T-4 YLIACGPMTK
T-2 placed at amino terminus because it begins with E (Glu).
T-3 placed at carboxyl terminus because it does not end with R (Arg) or K (Lys).
cleave with cyanogen bromide; separate fragments; sequence by Edman degradation
C-1 EGAAYHDFEPIDPRGASM
C-2 TKDCVHSD
C-3 ALIKYLIACGPM
C-3 overlaps with T-1 and T-4, allowing them to be ordered.
establish sequence
T-2
T-1
T-4
T-3
Amino terminus
EGAAYHDFEPIDPRGASMALIKYLIACGPMTKDCVHSD
Carboxyl terminus
C-1
C-3
C-2

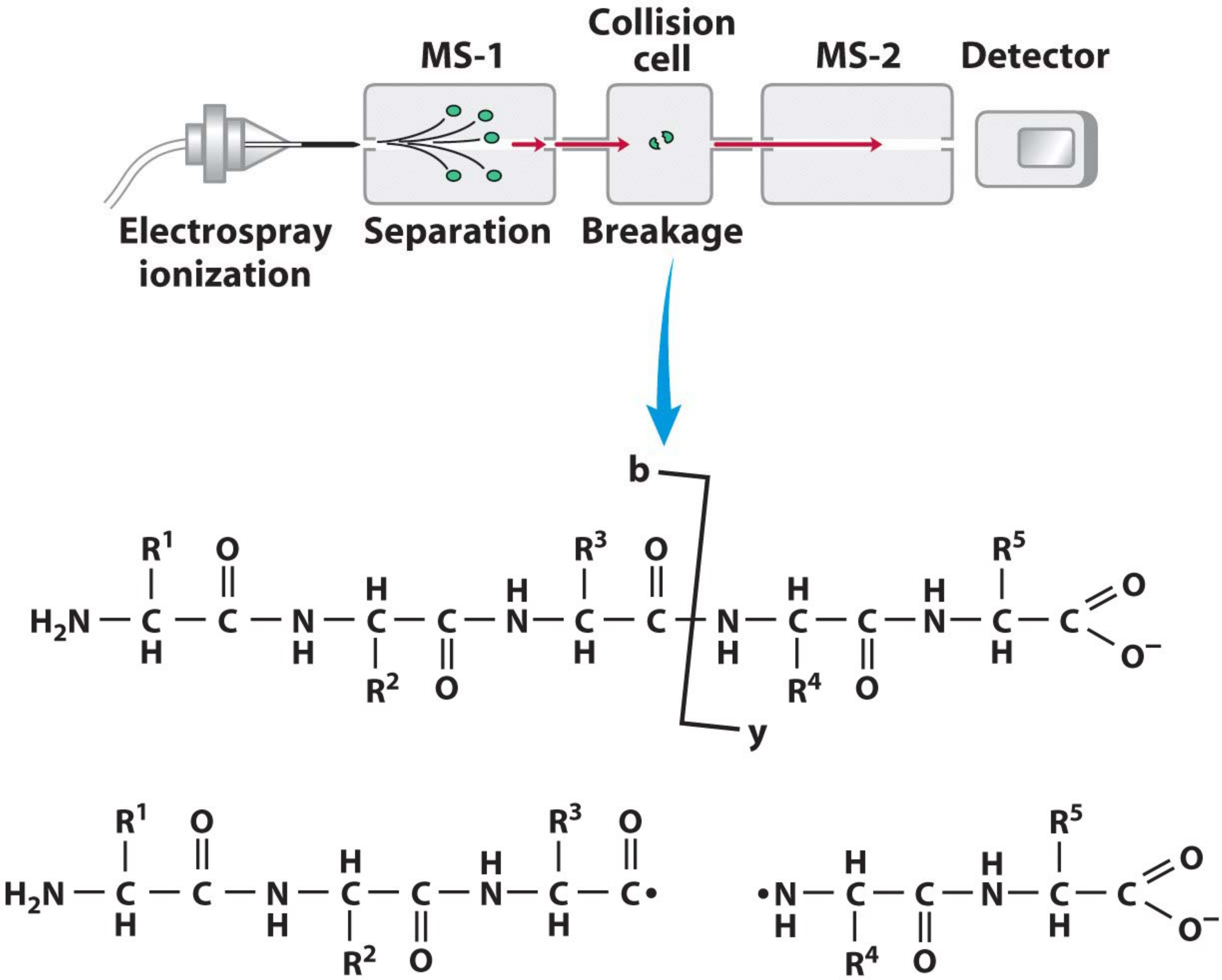
Amino acid sequence (protein) Gln – Tyr – Pro – Thr – Ile – Trp

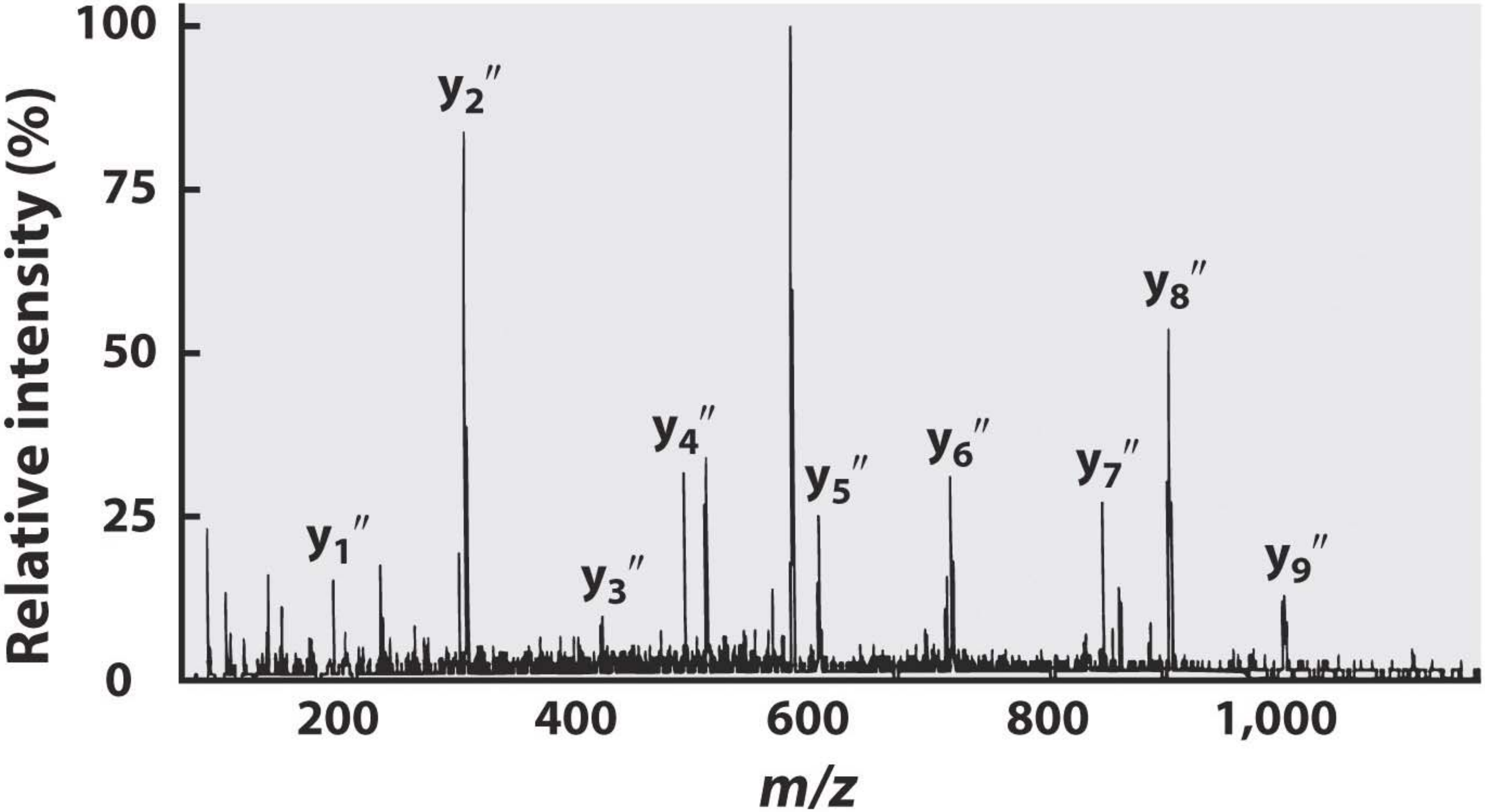
DNA sequence (gene) CAGTATCCTACGATTTCG


Glass
capillary
Sample
solution
High
voltage
Mass
spectrometer
Vacuum
interface

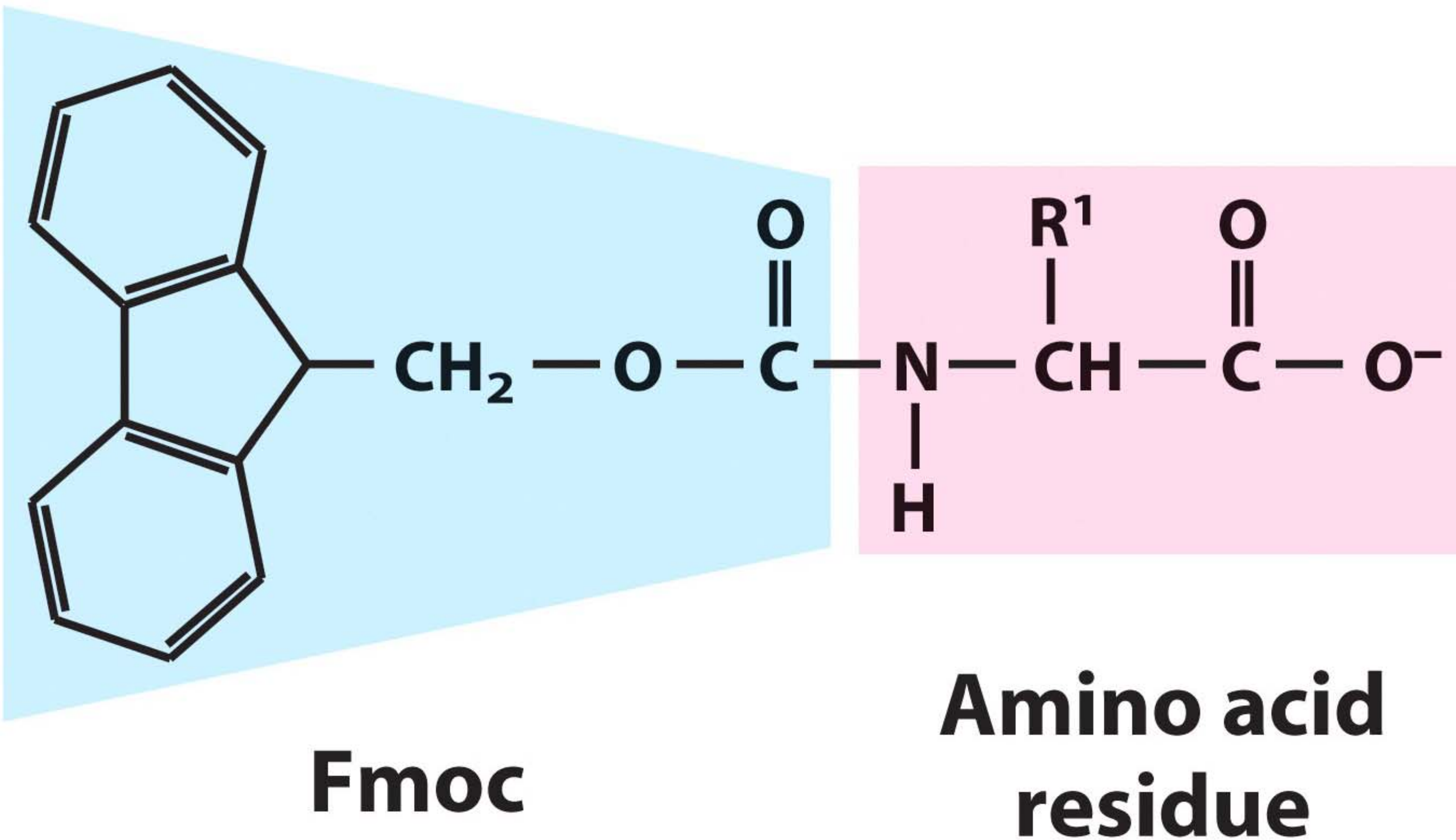


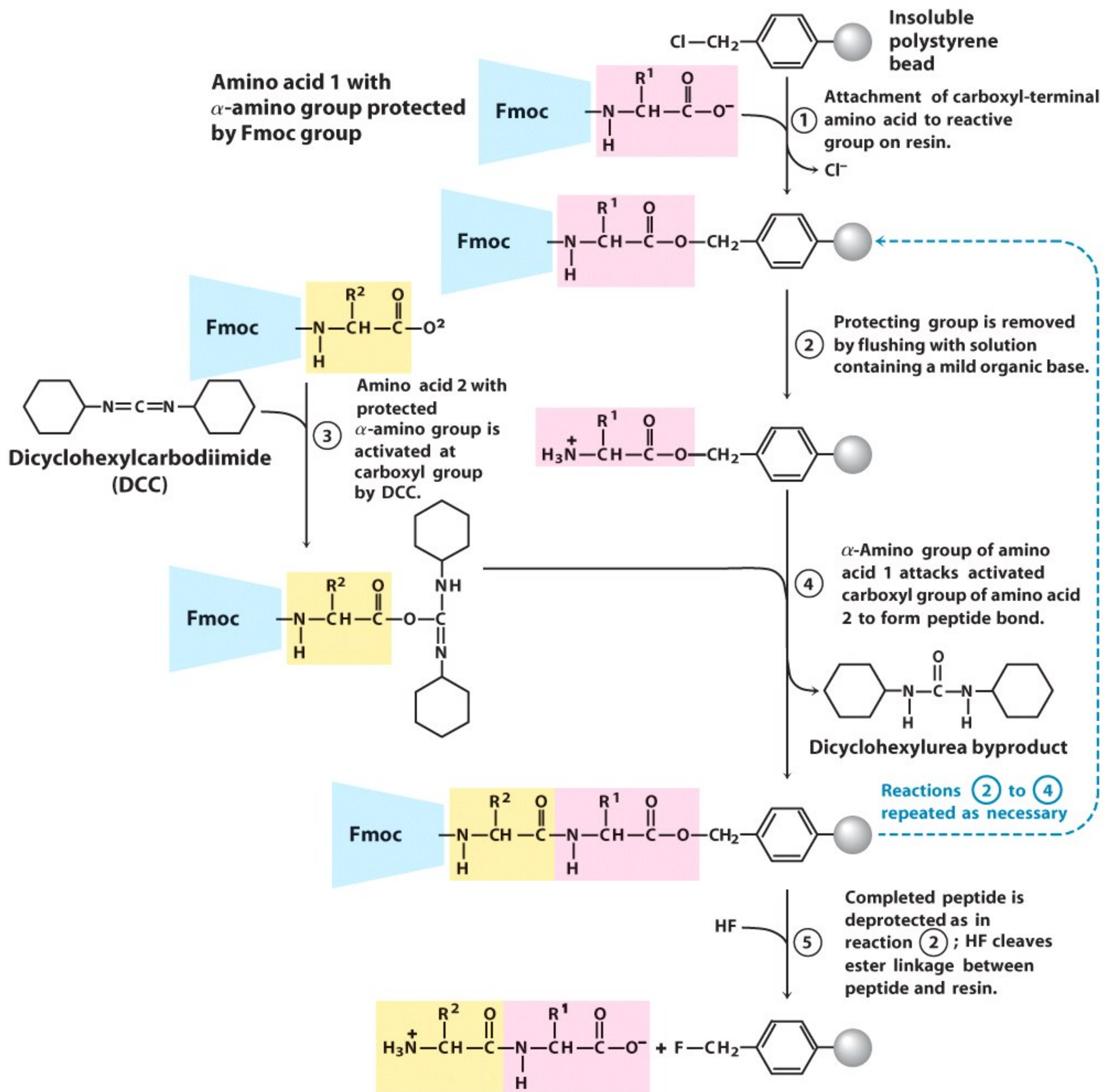
Relative intensity (%)
100
75
50
25
0
800
1,000
1,200
1,400
1,600
m/z
50+
40+
30+
47,342
100
50
0
47,000
48,000
M_r


Collision cell
MS-1
MS-2
Detector
Electrospray ionization
Separation
Breakage
b
y

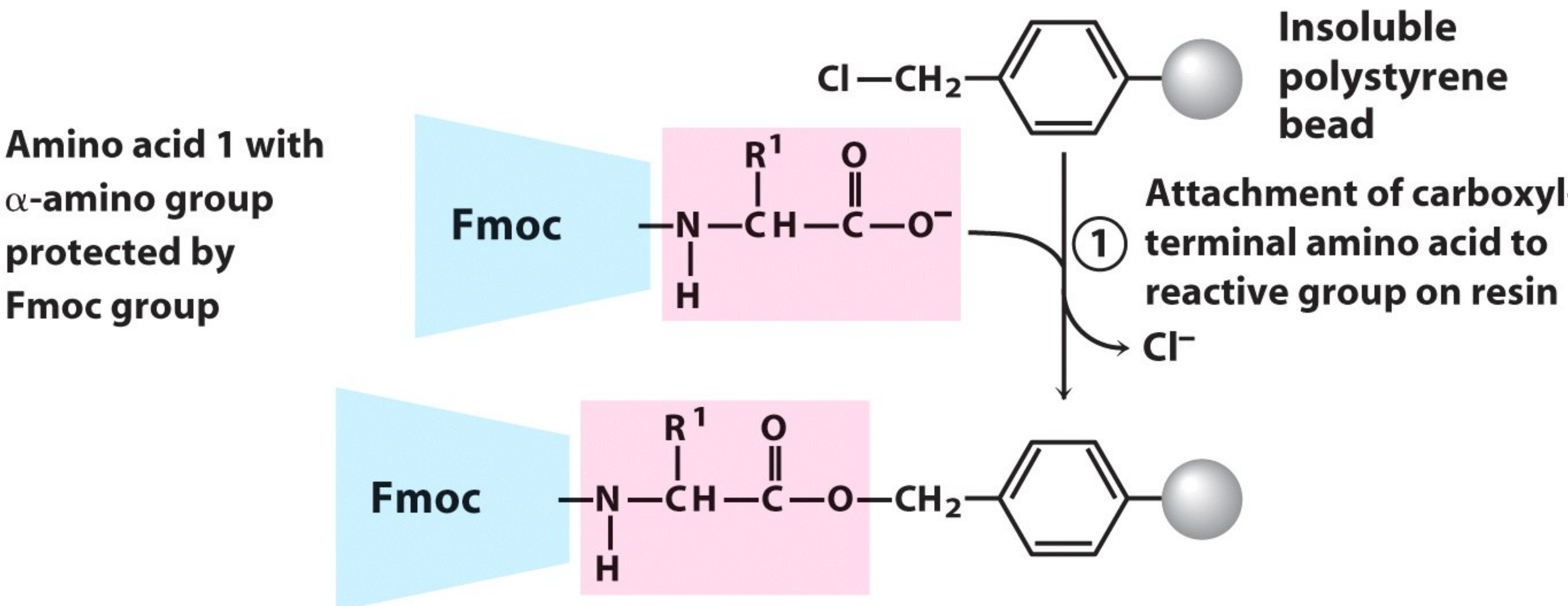


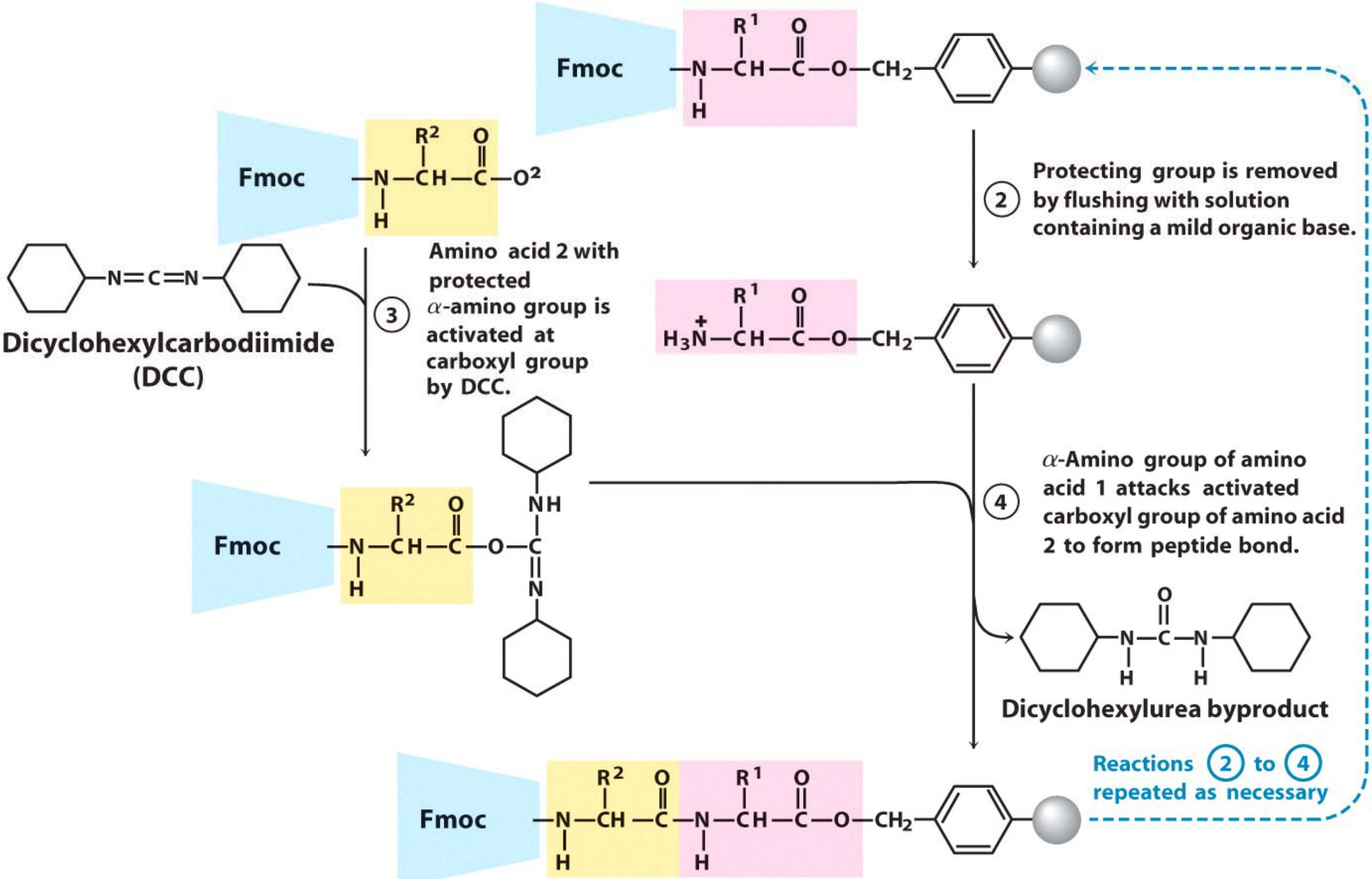
100
75
50
25
0
Relative intensity (%)
y2″
y1″
y3″
y4″
y5″
y6″
y7″
y8″
y9″
200
400
600
800
1,000
m/z


CH2
O
C
O
N
H
CH
R1
C
O
O−
Fmoc
Amino acid residue



Insoluble polystyrene bead
Amino acid 1 with α-amino group protected by Fmoc group
Fmoc
1 Attachment of carboxyl-terminal amino acid to reactive group on resin.
Cl⁻
2 Protecting group is removed by flushing with solution containing a mild organic base.
Dicyclohexylcarbodiimide (DCC)
3 Amino acid 2 with protected α-amino group is activated at carboxyl group by DCC.
4 α-Amino group of amino acid 1 attacks activated carboxyl group of amino acid 2 to form peptide bond.
Dicyclohexylurea byproduct
Reactions 2 to 4 repeated as necessary
HF
5 Completed peptide is deprotected as in reaction 2; HF cleaves ester linkage between peptide and resin.


Amino acid 1 with
α-amino group
protected by
Fmoc group
Insoluble
polystyrene
bead
Cl—CH2
Fmoc
R1
O
—N—CH—C—O−
H
1
Attachment of carboxyl-
terminal amino acid to
reactive group on resin
Cl−
Fmoc
R1
O
—N—CH—C—O—CH2
H


Fmoc
R1
O
N
CH
C
O
CH2
H
Fmoc
R2
O2
Protecting group is removed by flushing with solution containing a mild organic base.
2
N=C=N
Dicyclohexylcarbodiimide (DCC)
3
Amino acid 2 with protected α-amino group is activated at carboxyl group by DCC.
H3N
NH
α-Amino group of amino acid 1 attacks activated carboxyl group of amino acid 2 to form peptide bond.
4
Dicyclohexylurea byproduct
Reactions 2 to 4 repeated as necessary

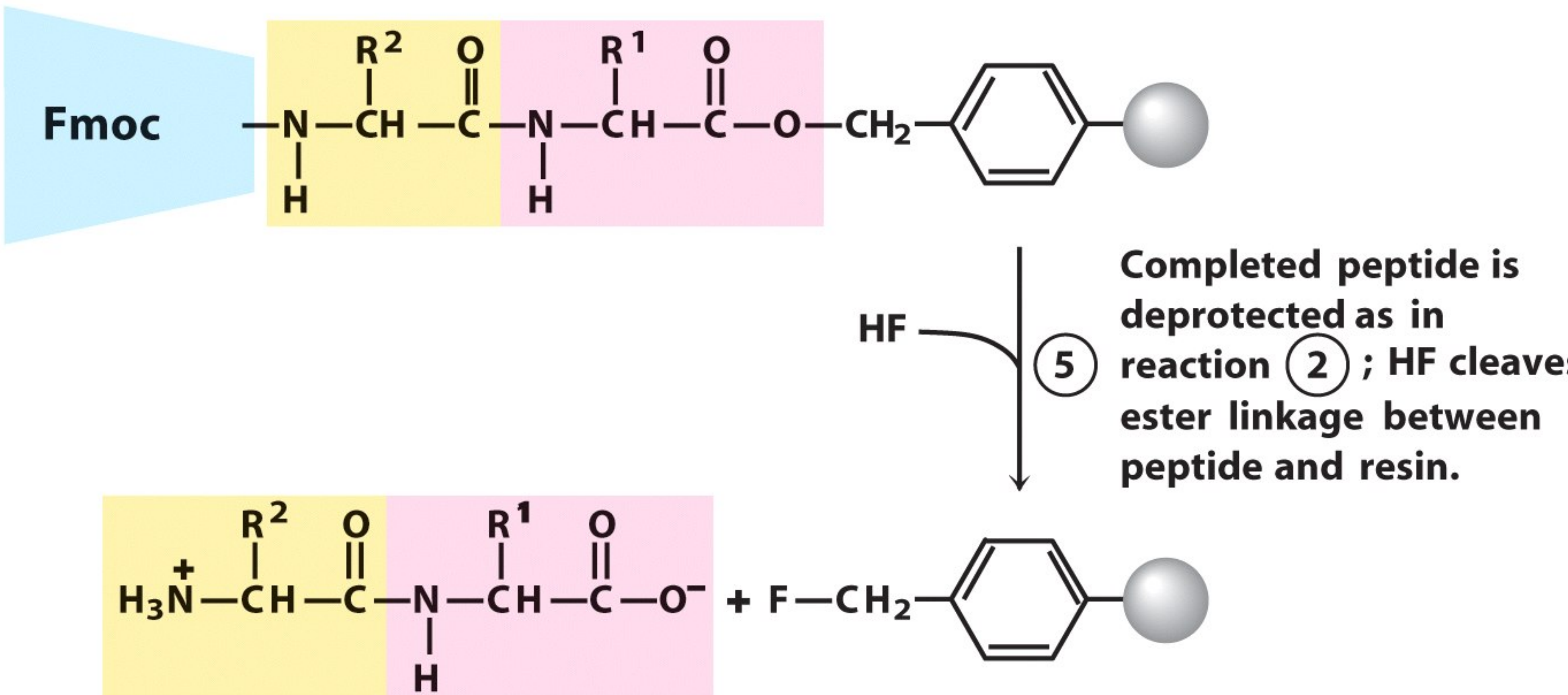

Fmoc
R2
O
N
H
CH
C
R1
O
C
O
CH2
HF
5
Completed peptide is deprotected as in reaction 2 ; HF cleaves ester linkage between peptide and resin.
H3N+
O−
+ F—CH2

TABLE 3–8 Effect of Stepwise Yield on Overall Yield in Peptide Synthesis

Number of residues in the final polypeptide	*Overall yield of final peptide (%) when the yield of each step is:*	
	96.0%	*99.8%*
11	66	98
21	44	96
31	29	94
51	13	90
100	1.7	82